AN ANALYSIS OF TEST ITEMS OF THE ENGLISH FIRST-TERM TEST OF THE SEVENTH GRADE STUDENTS OF SMP MUHAMMADIYAH 10 YOGYAKARTA IN THE ACADEMIC YEAR OF 2013/ 2014

Presented as Partial Fulfillment of the Requirements to Attain the Degree of
Sarjana Pendidikan in the English Education Department

by
Yunita Risydah
06202244051

ENGLISH EDUCATION DEPARTMENT
FACULTY OF LANGUAGES AND ARTS
YOGYAKARTA STATE UNIVERSITY
2014

AN ANALYSIS OF TEST ITEMS OF THE ENGLISH FIRST-TERM TEST OF THE SEVENTH GRADE STUDENTS OF SMP MUHAMMADIYAH 10 YOGYAKARTA IN THE ACADEMIC YEAR OF 2013/ 2014

Presented as Partial Fulfillment of the Requirements to Attain the Degree of *Sarjana Pendidikan* in the English Education Department

by

Yunita Risydah

06202244051

ENGLISH EDUCATION DEPARTMENT

FACULTY OF LANGUAGES AND ARTS

YOGYAKARTA STATE UNIVERSITY

2014

Approval Sheet

An Analysis of Test Items of the English First-Term Test of the Seventh Grade Students of SMP Muhammadiyah 10 Yogyakarta in the Academic Year of 2013/ 2014

A Thesis

By

Yunita Risydah/ 06202244051

Approved by

First Consultant

Dr. Agus Widyantoro, M.Pd.

NIP. 19600308 198502 1 001

RATIFICATION

An Analysis of Test Items of the English First-Term Test of the Seventh Grade Students of SMP Muhammadiyah 10 Yogyakarta in the Academic Year of 2013/ 2014

A Thesis

Accepted by the Board of Examiners of the Faculty of Languages and Arts of Yogyakarta State University on January 20th, 2014 and Declared to have fulfilled the Requirements to Acquire a Sarjana Pendidikan Degree in English Education.

Board of Examiners		Signatures
Chairman	: Drs. Samsul Maarif, M.A	.....
Secretary	: Lusi Nurhayati, S.Pd., M.App.Ling	.....
First Examiner	: Dr. Margana, M.Hum., m.A	.....
Second Examiner	: Dr. Agus Widyantoro, M. Pd	.....

Yogyakarta, 22 Januari 2014

Faculty of Languages and Arts

Yogyakarta State University

Dean,

Prof. Dr. Zamzani

NIP. 19550505 198011 1 001

Pernyataan

Yang bertandatangan dibawah ini:

Nama	: Yunita Risydah
Nim	: 06202244051
Program Studi	: Pendidikan Bahasa Inggris
Fakultas	: Bahasa dan Seni
Judul Skripsi	: **An Analysis of Test Items of the English First-Term Test of the Seventh Grade Students of SMP Muhammadiyah 10 Yogyakarta in the Academic Year of 2013/ 2014**

Menyatakan bahwa karya ilmiah ini adalah hasil karya saya sendiri dan sepengetahuan saya tidak berisi materi yang ditulis oleh orang lain kecuali bagian-bagian tertentu yang saya ambil acuan dengan mengikuti tata cara dan etika penulisan karya ilmiah yang lazim

Apabila terbukti bahwa pernyataan saya ini tidak benar, sepenuhnya menjadi tanggung jawab saya.

Yogyakarta,

Yang membuat pernyataan

(Yunita Risydah)

DEDICATIONS

In the name of Allah, I dedicate my thesis for:

My beloved father for his prayer, support and endless love, I love you so much

My late mother in heaven, I believe you always pray for me

My sisters and brother for the support, I love you all

All of Furqoners for their support, help and pray for me

All of my friends who have supported and helped me

MOTTOS

Do the best, let GOD do the rest!

(Anonymous)

"(Allah) Thee only do I worship, and Thee only do I ask for help.(Al-Fatihah:5)"

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful, all praise be to Allah for the strengths and His blessing for me in completing this thesis.

I would like to express my gratitude to my first consultant, Dr. Agus Widyantoro, M.Pd who has been very helpful in guiding and supervising me during the process of writing this thesis. My gratitude also goes to all of the lecturers of the English Education Department for their kindness during my study.

I would like to express my appreciation to the big family of SMP 10 Muhammadiyah Yogyakarta who have given me a permission to conduct the research, particularly to Bu Arifiana Budiarti and the students of class VII.

My deepest gratitude goes to my beloved father, sisters, and brother for their endless support, patient, prayers, and love. In addition, I'm also very grateful for the encouragement and support from my friends in Furqon boarding house who has supported and helped me so that I can finish writing this thesis. I also thank all my PBI'06 friends for their support and love.

Last but not least, I'm indebted to those who indirectly contributed in the accomplishment of this research. Their kindness means a lot to me.

Yogyakarta, Januari 2014

Penulis

TABLE OF CONTENT

LIST OF TABLES

LIST OF PICTURE

LIST OF APPENDICES

AN ANALYSIS OF TEST ITEMS OF THE ENGLISH FIRST-TERM TEST OF THE SEVENTH GRADE STUDENTS OF SMP MUHAMMADIYAH 10 YOGYAKARTA IN THE ACADEMIC YEAR OF 2013/ 2014

By
Yunita Risydah
06202244051

ABSTRACT

This study is aimed at finding out the quality of the test item of the first – term test of the seventh grade students SMP Muhammadiyah 10 Yogyakarta in the academic year 2013/2014 which consists of 45 objective tests and 5 essay tests. The test item is analyzed based on the validity, reliability, difficulty index, and discrimination index.

This research is classified into quantitative research. The data was collected from the test item, the key answer and the students' answer sheet which are analyzed using *Item and Test Analysis (ITEMAN)*. The participants of this study were the seventh grade students of SMP Muhammadiyah 10 Yogyakarta.

The results of the study show that: (1) based on the validity analysis, there are 75,56% of the test item which are valid, and 24,44% which are invalid, (2) reviewed from the reliability, the test item has reliability index 0,668 that is categorized as low reliability, (3) from the difficulty index, there are 28,89% items that difficult, 53,33% medium difficult, and 17,78% that categorized as easy, and (4) from the discrimination index analysis, there are 15,56% that have poor discrimation index, 17,78% have satisfactory discrimination index,24,44% have good discrimination index, and 42,22% that have excellent discrimination index (5) while from the distractor efficiency, there are 87% of the test items have excellent distractor efficiency and 13% which have good enough distractor efficiency. From the finding above, it can be concluded that the quality of the test items of English first term test of the seventh grade students of SMP Muhammadiyah 10 Yogyakarta is poor.

CHAPTER I

INTRODUCTION

A. Background of the Study

English language becomes one of the subjects of National Examination in junior high school and senior high school. In junior high school national examination, there are 4 subjects that being tested. They are Mathematics, Indonesian Language, Science and English. In this national examination, there is a passing grade. Nowadays, the passing grade is 4,0 as the minimum mark, and they must have 5,5 as the average mark of that four subjects. It means the students have to pass that passing grade, if they get lower mark they will fail.

The condition above causes the English teaching and learning activities complex, as we know that English is a foreign language for the students. So, it needs some effort in order to help the students achieve that goal. All teachers hope that all of their students can graduate with a good mark.

To know how far the students' mastery of English language is, the teacher usually uses certain measurements. It can be done by performing evaluation. Evaluation may be defined as a systematic process of determining the extent to which the instructional objectives are achieved by the pupils (Groundlund, 1981). He says that the evaluation results are used for describing the changes in their performance as well as value judgment. Therefore, the results show the students' learning progress.

When the teacher wants to know the students' progress in learning, they evaluate them. They need instrument to evaluate the achievement, which is called a test. A test is a particular type of assessment that typically consists of a set of questions administered during a fixed period of time under reasonably comparable conditions for all students (Miller, 2009).

The goal of the testing should be in accordance with the material that has been taught. There are various kinds of test which is done by the teachers in order to measure their students' knowledge, ability, or performance of English itself. We can see there are: midterm test, term test, and national examination. It is aimed to know how far the students mastery of the material given. There is a minimum passing grade which is usually called as Kriteria Ketuntasan Minimal (KKM). Students who get lower grade than KKM have to face remedial test. KKM might be different in each school. By doing some tests above, the teacher knows how far the success of the teaching and learning process. It means that a test is an evaluation instrument to measure the students' ability, knowledge, or performance of English.

In SMP 10 Muhammadiyah Yogyakarta, the first term tests are developed by team. The members of this team are teachers of the same subjects from different Muhammadiyah School in Yogyakarta province, which is usually called as MGMP.

The test results got by the students are varied. In reality we can see that there are some students found difficulties in doing English test. It happened because of some factors. They are the students' knowledge of the subject, the way

of teaching and learning process given by the teacher, the quality of the test itself, etc.

The quality of good test can be affected by some factors. They are validity, reliability, discrimination index, the difficulty index, and the effectiveness of distractor. Besides, a good test must reflect a conclusion of the objectives; it must have high readability and there is enough time to do the test.

In relation to the criteria above, an analysis of the test items is a way to improve the quality of the evaluation instrument. Therefore, teachers and the test designers should have ability to analyze the test item. But, based on the observation, the English teachers in SMP Muhammadiyah 10 Yogyakarta knew whether a test was good or not based on the students' answer and how many students who could answer the test correctly, not from analyzing the test item. Teacher and the test designer tend to use the preceding test material to make a new test. It make the quality of test material is not yet known from the validity, reliability, difficulty index, discrimination index and efficiency distractor. A quality of a good test material can be discovered by analyze the test item, so the test designer can decides which item should be kept, which item should be revised, and which item should be omitted.

From the illustration above, the writer state the reasons for choosing the title is to find out the quality of the test item of the English first term for the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014. This study is useful to improve the evaluation instrument, especially the English evaluation instrument.

In this study, the writer analyzed the quality of the test item using quantitative methods using *Item and Test Analysis (ITEMAN)*. There are five characteristics that usually determined for a test item: (1) validity; (2) reliability; (3) item difficulty; (4) item discrimination; and (5) Distractor Analysis.

B. Identification of the Problem

After analyzing the problems found in the field previously, the researcher then collected some problems related to quality of the test item. These problems were eliminated and chosen in limitation of the problem and discussed in this research in the next part. The problems are as follows:

1. The teacher did not analyze the test material given to the students in the English first term test yet.
2. The teacher knew whether a test item was good or not from the students' answer and how many students can answer the test correctly.
3. The quality of the English test item of the first term test of SMP Muhammadiyah 10 Yogyakarta was not known yet.

C. Limitation of the Problem

Due to the limitation of time, the researcher only chose one problem to be studied. Based on the background of the problem that has been explained above, the problem of this research is limited to find out the quality of the English first term test items of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014. Moreover, because of the writer's

limitation of knowledge, time and funds, the researcher only analyzes the objective tests.

D. Formulation of the Problem

Based on the limitation problem above, the formulation of the problem in this research is:

How is the quality of the English first term test items of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014?

E. Objectives of the Research

In accordance with the formulation of the problem above, the objective of this research is to find out the quality of the English first term test items of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014.

F. Significances of the Research

The findings are expected to give contribution to:

1. Teacher and students of English department

   This research is useful for the teacher in general, and particularly for English teacher in SMP Muhammadiyah 10 Yogyakarta. He/she can evaluate their teaching and learning process, whether the materials

given are suitable with the curriculum or not. And measure their success on teaching.

This research is also useful for the students of English department. It gives the students of English department knowledge about a good test, how to make a good test, and how to analyze a test.

2. English test item developers

   This research shows how far the quality of a good test based on some criteria. Then, the result can be used as a feed-back for the test developer in order that they can improve the quality of the next test items construction.

3. Further research development.

   The findings of this research are expected to be a starting point for other researchers in conducting research especially on English evaluation which is still rare.

CHAPTER II

LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

A. Literature Review

In this part, the writer presents theories related to the topic discussed in the study. They are the first term test of Junior High School, the criteria of a good test and the item test analysis.

1. First –Term Test

A school conducts twice term tests in an academic year; they are first term test and second term test. First term test is done in the end of the first term. The test is called Quality Assurance Test (Tes Kendali Mutu). This test is done in order to know the students' mastery of the material given by the teacher.

There are some definitions of tests. First, a test is a systematic and objective tool or procedure to collect data which is systematic and objective about somebody in an accurate way (Kusuma in Arikunto, 2001:32). Secondly, a test is any series of questions or exercises or other means of measuring skills, knowledge, intelligence, or capabilities (Anderson, 1976:425)

According to Brown (2004: 3), a test is a method of measuring a person's ability, knowledge or performance in a given domain. A test is a method. It is an instrument, a set of techniques, procedure, or items that requires performance of the test taker. To qualify as a test, the method must be explicit and structured. A test must measure. Some tests measure general ability, while others focus on very specific competencies or objectives. A multi-skill proficiency test determines a

general ability level; a quiz on recognizing correct use of the define articles measures specific knowledge. That is way the results or measurements are communicated may vary.

A well-constructed test must accurately measure individual's ability within particular domain. It is important to note that each test taker may have different background and learning experience. To measure individual's ability accurately, testers need to understand who the test-takers are.

Brown (2004: 4) stated that tests are prepared administrative procedures that occur at identifiable times in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated. It is meant that the teacher has told to the student before he will give a test in the meeting; or teacher has to make an agreement with the students that they will have a test when they have finished discuss one topic or at the end of the semester.

Based on the theories above, the writer concludes that a test is an instrument. It is systematic and as objective tool or procedure to collect data that consists of any series of questions or exercises or other means of measuring skills, knowledge, intelligence, or capabilities. The first-term test is an instrument that is used by the teacher in order to measure the students' skill, knowledge, and their capabilities of English.

A test which is conducted at the end of a course or unit of instruction is called summative test or summative assessment. It aims to measure or summarize what a student has grasped. A summative test is done after the teaching and

learning activities or all of the planning programs have been completed. This test is done in to measure the level of the students' success to achieve the objectives of the topic. The result of this test can be used by the teacher to make consideration whether a teaching can be continued or should be repeated. Information that is obtained from a summative test can be used to decide the grade that is achieved by a student.

However, in a wider domain, teacher can assess students in the process of "forming" their competencies and skills. It is called formative assessment (Brown, 2004: 6). Besides, a teacher may give a test before he gives the material. It is usually called pre-test. This test is used to know what the students have known and what they need to know. So, the teacher can give them suitable material with appropriate method. A formative test is done during a teaching and learning activity, at the end of every topic. Formative tests are done many times in a semester. The aim of this test is to measure the level of the students' success to achieve the objectives of the topic. Information that is obtained from the test is an input to assess the effectiveness of the teaching and learning activities. For the teacher, for example, such information can be used to make consideration whether a teaching can be continued, whether it needs to be repeated for certain parts, or whether the teaching technique that is used must be changed, modified, or can still be used. If the students have not mastered the teaching material that is requirement to the next material, the teacher has to repeat the presentation of the material. If it is necessary, the repetition is done by changing the teaching technique and adding some more necessary material. For the students, formative

test give some advantages. First, students will understand whether they have mastered the material of certain topic they learned. Secondly, the results of the formative test can become reinforcement. If the result is good, the students will know that their answers are correct and then they will retain them in their possession. But if the result is poor, they will study harder to find the right answers.

Based on the explanation above, the first - term test is classified into summative test. It aims to measure or summarize the material that a student has grasped. The result that is obtained from this first-term test can be used to decide the grade that is achieved by a student. This information can also be used to know a student's position among his/her friends. Secondly, it is can be used to decide whether a student can be promoted into the next grade or not.

2. Form of Test

There are spoken and written tests in language program which have some types of tests. In written test, there are essay-type tests and objective-type tests. An essay-type test can be in the form of composition, dictation, or translation. Objective tests can be in the form of true-false, multiple-choice, matching, or completion.

a. Essay-Type Test

An essay test contains tasks that require the testees to write sentences or paragraphs in a certain language. The testees need to think about and use what they know about the answer to the question. An essay test gives freedom to the students

to arrange and suggest their own answer in some limited area (Tuckman in Nurgiantoro, 1987:11). Because of that, it is often called as a subjective test.

Subjective test give a chance to the testees to show their ability, apply knowledge, analyze, correlate, or evaluate information. They may work out facts and concepts and organize them into logical coherency, and then put down their thinking in a writing form (Ebel in Nurgiyanto, 1987:68)

Testees' answers in taking an essay test show the quality of their thinking, cognitive activity in a high degree that is not only about remembering or understanding. In evaluation, the way testees think and what is concluded are not important. The most important thing consists of the facts the testees present and the reason that lead to the conclusion. So the most important thing is not how to conclude, but how to make conclusion.

An essay test is best to apply in small classes, where the instrument is not used anymore; to make students dare to suggest their thinking ability in the high degree of the cognitive skill; to evaluate students' thinking process; to make teachers sure that they have the ability to act as critical reader; and to make teachers sure that they have enough time to correct the students' work.

b. Objective-Type Test

In the objective test, students have to choose certain codes to represent the alternatives of the required answers. The answer of an objective test is something exact. There is only one true answer. Because the answers are something exact, whoever becomes the scorer will result the same score.

There are some kinds of objective test. A true-false test is a test form that contains a statement that has a possibility of being true or false. Testees must understand the statements that are given to them. Then, if they think that the statement is right, they answer “yes”, or “true”. But, if they think that the statement is wrong, they answer “no”, or “false”.

Multiple- choice tests consist of a statement about a definition that is not complete. To complete the definition, the testees choose one of the possible answers that are prepared. So, a multiple-choice test consists of a stem and a number of alternative options. The alternative options consist of one true answer and some distractors.

A matching test is usually a series of comparisons or puzzles. A matching test consists of a series of questions and answers. Every question has an answer in a serial form. Testees have to find and place the answers so that they match with the questions.

Finally, a completion test consists of sentences with some parts deleted. Students must fill the parts that are deleted. The missing part can be a word, a phrase, or a clause.

3. Language Assessing

a. Assessing Listening

Every teacher of language knows that one’s oral production ability- other than monologues, speeches, reading aloud, and then like- is only as good as one’s listening comprehension ability. We need to pay close attention to listening as a mode of performance for assessment in the classroom.

There are four commonly identified types of listening performance, each of which comprises a category within which to consider assessment task and procedures. They are intensive, responsive, selective, and extensive type of listening.

The example of authentic extensive listening is found in a popular genre of assessment task in which the test-taker is presented with a stimulus monologue or conversation and then asked to respond to a set of comprehension question.

Here is the typical example of extensive listening:

Test-takers hear:

Doctor : Good morning, Lynn. What's the problem?
Lynn : Well, you see, I have a terrible headache; my nose is running, and I 'm really dizzy.
Doctor : Anything else?
Lynn : I've been coughing, I think I have a fever, and my stomach aches.
Doctor : I see. When did it start?
Lynn : Well, let's see. I went to the lake last weekend, and after I returned home, I started sneezing.
Doctor : hmmm, you must have the flu. You should get lots of rest, drink hot beverages, and stay warms. Do you follow me?
Lynn : Well, uh, yeah, but…. Shouldn't I take some medicine?
Doctor : Sleep and rest are as good as medicine when you have the flu.
Lynn : Okay, thanks, Dr. Brown.

Test-takers read:

1. What is Lynn's problem?
   a. She feels horrible
   b. She ran too fast at the lake
   c. She's been drinking too many hot beverages
2. When did Lynn's problem start?
   a. When she saw her doctor
   b. Before she went to the lake
   c. After she came home from the lake

Taken from: Language Assessment book page 133

b. Assessing Reading

In reading variety of performance is derived more from multiplicity of types of texts than from the variety of overt types of performance. Nevertheless, for considering assessment procedures, several types of reading performance are typically identified, and these will serve as organizers of various assessment tasks. The types of reading performance are: 1. Perceptive, 2. Selective, 3. Interactive, 4. extensive.

1) Perceptive. Perceptive reading tasks involve attending to the components of larger stretches of discourse: letters, words, punctuation, and other graphemic symbols. Bottom-up processing is implied.

2) Selective. This category is largely an artifact of assessment formats. In order to ascertain ones reading recognition of lexical, grammatical, or discourse features of language within a very short scratch of language, certain typical tasks are used: picture-cued tasks, matching, true/false, multiple-choice, etc.

3) Interactive. Included among interactive reading types are stretches of language of several paragraph to one page or more in which the reader must interact with the text. That is reading a process of negotiating meaning; the reader brings to the text a set of schemata for understanding it, and intake is the product of that interaction. Typical genres that lend themselves to interactive reading are anecdotes, short narratives, and descriptions, excerts from longer texts, questionnaires, memos, announcements, directions, recipes, and the like. The focus of an interactive task is to identify relevant features (lexical, symbolic,

grammatical, and discourse) within texts of moderately short length with the objective of retaining the information that is processed.

In teaching and learning of English in vocational school the selective reading type is used. The test items are in the form of impromptu reading plus comprehension, editing task, selected response fill- in vocabulary task and multiple-choice cloze vocabulary/grammar task.

Here is the typical example of selective reading type:

Multiple-choice cloze vocabulary/ grammar task:

He showed his suitcase (1)…..me, but it wasn't big (2)….to fit all his clothes. So I gave him my suitcase, which was (3)…….

1. A. for
   B. to
   C. from
2. A. so
   B. too
   C. enough
3. A. larger
   B. smaller
   C. largest

Taken from: Language Assessment book page 196

c. Assessing Writing

Four categories of written performance that capture the range of written production are considered here. Each category resembles the categories defined for the three skills, but these categories, as always, reflect the unique of the skill area. The categories are 1. Imitative, 2. Intensive (controlled) 3. Responsive, and 4. Extensive.

1) Imitative: To produce written language, the learner must attain skills in the fundamental, basic tasks of writing letters, words punctuation, and very brief sentences. This category includes the ability to spell

correctly and to perceive phoneme-grapheme correspondences in the English spelling system. At this stage, form is the primary if not exclusive focus, while context and meaning are of secondary concern.

2) Intensive (controlled). Beyond the fundamentals of imitative writing are skills in producing appropriate vocabulary within a context, collocations and idioms, and correct grammatical features up to the length of a sentence. Meaning and context are of some importance in determining correctness and appropriateness, but most assessment tasks are more concerned with a focus on form, and are rather strictly controlled by the test design.
3) Responsive. In this category, assessment tasks require learners to perform at a limited discourse level, connecting sentences into paragraph and creating a logically connected sequence of two or three paragraphs. Tasks respond to pedagogical directives, list of criteria, outline, and other guidelines. Genres of writing include brief narratives and descriptions, short reports, lab reports, summaries, brief responses to reading, and interpretation of charts or graph. Form-focused attention is mostly at discourse level with a strong emphasis on context and meaning.
4) Extensive. Extensive writing implies successful management of all the processes and strategies of writing for all purposes, up to the length of an essay, a term paper, a major research project report, or even a thesis. Writers focus on achieving a purpose, organizing and

> developing ideas logically, using detail to support or illustrate ideas, demonstrating syntactic and lexical variety, and in many cases, engaging in the process of multiple drafts to achieve a final product. Focus on grammatical form is limited to occasional editing or proofreading of a draft.

The type of writing assessment usually used as the practice test of writing by students. In Junior High School, it is done when the students are in the third grade. They have to create a short text based on certain genre.

4. Characteristics of a Good Test

Arikunto (2001:11) states there are several characteristics of an educational test. First, the test measures the skill indirectly. For example, when it is used to measure intelligence, it tests the ability to answer the test items. Second, the test uses quantitative measures. It uses numeric symbols as the first results of measurement. Third, it is interpreted in the qualitative form. Forth, the test uses consistent units. For example, in an IQ test, the numbers of 0 to 30 is used to show whether a testee is idiot or a genius. Fifth, the test has relative characteristics. It means that the results may be different from time to time. For example, a testee may score 80 on a test on Monday, 90 on Wednesday, and 70 on Friday.

There are some characteristics of a good test. Arikunto (2001) states that a good test must be valid, reliable, and have a normal discrimination index, difficulty index and a good efficiency distractor.

1. Validity

The first and foremost question to be asked with respect in any testing procedure is: how valid is it? When this question is asked, it inquires whether the test measure what the people want to measure, all of what to people to measure, and nothing but what the people want to measure.

The validity of the score is not self-evident but is something that must be established on the basis of adequate evidence. A test may be thought of as corresponding to some aspect of human behavior in any one of the three senses. The terms that have been adopted to designate these senses are content validity, criterion validity, and construct validity.

a. Content Validity

If the measurement of English competency will be made, the first thing to do is reading some agreement as to the skill and knowledge that comprise correct and effective use of English, and that have been the objectives of language instructions. The next step is examining the test to see what skills, knowledge, and understanding it calls for. The final step is matching the analysis of test content against the analysis of course content and instructional objectives and seeing how well the former represents the latter.

It should be clear that relationship between teaching and testing is typically intimate. Test is drawn from what has been taught, or what is the purpose to be taught. The instructional program is the original source of test materials.

b. Criterion- Related Validity

Frequently, a test is used in correlation with decision that implies predicting some specific future outcome. Evaluation of a test as predicting is primarily an empirical and statistical evaluation, and this aspect of validity has some time been spoken as empirical or statistical validity. The basic procedure is to give the test to a group who are entering some job or training program, to follow them up later to get for each one specified criterion measure of success on the job or training program, and then to complete the correlation between test scores and the criterion measure of success. The higher the correlation, the more effective the test as predictor. If a test is analyzed using criterion-related validity, and the result is high, it means the test is effective as a predictor.

c. Construct Validity

Some questions that are asked in psychology tests are neither “how well does this test predict job success?” nor “How well does this test represent our curriculum?”, but rather “What do score on this test mean or sight?” “What does the score tell about an individual?” “Does it correspond to some meaningful trait or construct that will help in understanding him?” For this question of whether the test tells something meaningful about people, the term construct validity has been used.

Some educational tests are intended to measure general traits or qualities of an individual. Verbal reasoning, special visualizing, sociability, introversion, mechanical interests are all designation of traits and constructs. Tests of these

functions are valid in so far as they behave in the way that such a trait should reasonably be expected to behave.

The validity of the test can be found using formula of correlations coefficients. The formula of $^{y}_{pbi}$ (point-biserial correlation coefficients) is used. Brown (2005:162) states the appropriate statistic to apply (when examining the relationship between a nominal scale, like right or wrong, and a continuous scale) is the point-biserial correlation coefficients. Arikunto (2003:78) states besides using the product-moment correlation, there is other formula to calculate the validity of test item which is called point-biserial correlation coefficients. This formula calculates mean of correct answer and the mean of incorrect answer. The higher value of $^{y}_{pbi}$ the higher value of the test item validity. The categories of validity value are very high, high, medium, low, and very low.

2. Reliability

Reliability can be defined as how far a test result can stay the same in different objects and situations. There are some ways to calculate reliability; those are stability, equivalence, and internal constancy. Stability can be found through a test-retest technique. Equivalence can be found through the parallel method. Internal consistency can be found through split half methods, standard error of measurement, item response theory, Cronbach alpha methods, and Hyor methods.

a. Test-retest

This technique is used to know how precisely a given measurement characterizes a person from day to day- how his score next week could be predicted from what he does today. It would be appropriate to measure him on

two separate occasions. Variation of the individual from one time to another time as variation due to the operation of measurement will be happen. There are three main sources of variation in performance that will tend to reduce the precision of a particular score as a description of an individual: 1) variation from trial to trial in response to the task at a particular moment in time, 2) variation in the individual from one time to another, and 3) variation arising out of the particular sample of tasks chosen to represent a domain behavior.

b. Parallel Test Form

Parallel test forms consist of two tests that are arranged on different forms but according to a specification of a difficulty degree that is balanced. Students are faced to the two types of parallel test forms in the same time. Reliability of these two measurement forms is obtained by correlating the score results of these two measurements.

c. Split-half method

This approach is very similar to the equivalent-forms technique except that, in this case, the equivalent form are created from the single test will be analyzed by dividing it into two equal part (Brown, 2005:177) . A widely used procedure for doing this is by dividing the existing test up into two presumably equivalent halves. The half-test may be assembled on the basis of careful examination of the content and difficulty of each item, making a systematic effort to balance out the content and difficulty level of the two halves. A simpler procedure is to put alternative items into two half-tests; that is, to put all the odd numbered items in one half-test and all the even-numbered items in the other half

and correlation between these two sets of scores will result in degree of measurement accuracy. This approach is very similar to the equivalent-forms technique except that, in this case, the equivalent form are created from the single test will be analyzed by dividing it into two equal part.

To avoid the work complexity involved in the test-retest and equivalent forms strategies, testers most often use internal-consistency strategies to estimate the internal-consistency reliability. As the main implies, internal-consistency reliability strategies estimates the consistency of a test using only information internal to a test that is available in one administration of a single test (Brown: 2005: 176).

The strategy to understand internal-consistency of the test items is called split-half method (split-half reliability). This approach is equivalent-forms technique except that, in this case, the equivalent forms are created from the singe test being analyzed by dividing into two equal parts. The other alternatives to calculate the internal-consistency of test items is calculating the Cronbach alpha coefficient. It gives very similar result to the split-half method. The higher the split-half reliability coefficient, the more consistent are the items on the test (in the sense that odd numbered items measure the same variable as the even-numbered items).

3. Difficulty Index

A good test item is not easy and not so difficult either. The items that are very easy will not stimulate students to study hard. On the other hand, very difficult test items make students hopeless and not have the spirit to try again. A

difficulty index is a numeral that shows whether a test is difficult or easy. The index runs between 0.00 to 1.0. Test items with a difficulty index of 0.00 shows that the test is very difficult. And one with a difficulty index of 1.0 shows that the test is very easy.

0.0 ——————— 1.0

Difficult Easy

4. Discrimination Index

Discrimination is the ability of the test items to discriminate clever from poor students. The numeral shows the discrimination is the discrimination index. A discrimination index is between 0.0 and 1.0. All the testees are divided into the clever or upper group and the poor or lower group. If the entire upper group's answers are true and all of the lower group's answers are false, the discrimination index is 1.00. If all of the upper group's answers are false and the lower group's answers are true, the discrimination index is – 1.0. And if all of the upper and lower group's answers are the same (false or true), the discrimination index is 0.

According to Arikunto (2001:2:218), discrimination index is classified into:

Discrimination (D) = 0.00- 0.2: poor

Discrimination (D) = 0.20- 0.40: satisfactory

Discrimination (D) = 0.40- 0.70: Good

Discrimination (D) = 0.70- 1.00: Excellent.

Negative discrimination index; (-D) the entire negative discrimination index is poor.

5. Distractor efficiency

The efficiency of distractors is the extent to which (a) the distractors "lure" a sufficient number of test-takers especially lower ability ones, and (b) those responses are somewhat evenly distributed across all distractor (Brown, 2004: 60).

The efficiency of the distractor can be seen by looking at the distribution pattern of the students' answer. The distribution pattern answer can be obtained by counting the test-takers who choose the alternative answer or the test takers who do not choose anything. From the distribution pattern answer, it can be determined whether the distractor is efficient or not.

5. Item Test Analysis

The analysis of the test items is an evaluation effort to get tests with good quality and relevancy criterion. A test item analysis needs to show which items are very difficult and do not function normally. The main goal of the test item analysis is to determine the difficulty index and discrimination index of every item of a test. If the analysis of test items has been done, one will understand the thinking process and thinking degree of students and their actions when they do the test. According to Brown (2004: 41), item analysis is the systematic evaluation of the effectiveness of the individual items on a test. This is usually done for purposes of selecting the best item which will remain on a revised and improved version of the test.

The test item analysis will also give the skills to the teacher (instructor) in test development. The test item analysis improves not only the individual quality, but also the quality of an educational technique. The aims of the analysis, according to Throndike and Hagen in Purwanto (1991: 119), are: to search the lesson of the class and the learning failure and then to guide the students to the better learning way and to prepare better tests for the following years.

So, the aims of the test item analysis are to look for which items are good and bad and why they are good or bad. By identifying the test item analysis, the factors which cause the items to be bad can be found.

B. CONCEPTUAL FRAMEWORK

Test has an important role in teaching learning process. It is impossible for teacher to assess the teaching achievement and progress without conducting any assessment or test. A good test must be able to measure the student's ability, performance and knowledge accurately.

First - term test of Junior High Schools are constructed by the team that consist of some teachers in the same subject from different schools. It means the test is constructed based on the teaching learning process and the perception of the teacher on the ability of the students.

In constructing a test, teacher should pay attention to the principles of constructing a good test. There are at least four principles; they are reliability, reliability, difficulty index, and the discrimination index (Brown, 2004). The

writer intends to find out the four criteria based on the score got by students in the even - term test using the appropriate instruments.

It is important to analyze the items of the test. By analyzing the test items, teacher knows whether the test is categorized as a good test or not. Teacher can learn how it forms and develop his own test. Even, teacher can avoid constructing unqualified test and construct a better one.

CHAPTER III

RESEARCH METHOD

A. Research Design

This research is aimed to find out the quality of the test items based on the score got by the students. First of all, the writer collected the data, they are in the form of test item and students' answer sheets, then did scoring to the data. After that, the researcher calculated the data using the suitable instruments. The result of this calculation is used to find out the measurements of the five criteria of a good test. In this study the writer made the interpretation based on the data in the field. There is no writer's objectives interruption in making the interpretation. It is a neutral interpretation based on the real data and its value category.

Based on its characteristic, this research is categorized as a quantitative research. It's the systematic science investigation of quantitative properties and phenomena and their relationships. The objective of quantitative research is to develop and employ mathematical models, theories, and or hypothesis pertaining to natural phenomena. Quantitative research in education has, thus, attempted to discover existing facts under the research belief that the research act must be a neutral activity from the researcher's subjective viewpoint (Smith, 1983). Thus, Smith (1983) places quantitative research as a "journey of the facts" (p. 10).

According to Smith, there was a belief that "neutral, scientific language" (p.9) must be used in quantitative research in order to find out exact facts. It was believed that using neutral scientific language was effective not only for providing

the research facts but also for explaining the statistical truth. In addition, neutral scientific language was able to directly show the results of research without a researcher's value judgments. That is to say, the research results existed in isolation from the researcher's viewpoint. According to Carr and Kemmis (1986:89), a researcher was considered to be "an outsider to the research" or an objective observer.

In this study the writer did a research to the test items of the first term English test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta. The writer used quantitative research method to analyze the data. As stated in the characteristic of a quantitative research above, the writer developed the instrument and defined measurement based on the instrument itself. Then, the writer made interpretation of the data without having any subjective viewpoint. It is totally based on the data in the context.

B. Place And Time of Study

In this analysis, the researcher analyzed a written data. The data were taken from one of Junior High school in Yogyakarta. The school is SMP 10 Muhammadiyah Yogyakarta. The research was conducted from December 2013 up to January 2014.

C. Subject of the Study

The subject of this study is the first-term test of the seventh grade students in SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014. The

first - term test items are in the form of multiple choices. This test was conducted in the same level with the same number of items and time limitation.

Table 1. Subject of the study

Class	**Number of Students**
VII A	37
VII B	37
VII C	36
Total	**110**

This test was conducted on Wednesday, the 11st December 2013. There were 50 items. The test consists of 45 objective-type tests and 5 easy-type tests. Each of the objective type items has 4 options (A, B, C, and D). There was a time limitation in doing this test. The students should finish the test in 2 hours, started from 07.30- 09.30 am. In scoring the students result, each of the items that were answered correctly was score or point 1. The total number of score is 70. In dividing the final score, the total score was divided by 7.

D. Population and Sample

According to Arikunta (2002:108), a population is a set (or a collection) of all elements processing one or more attributes of interest. Then Arikunto (2002:112) also states that "if the population is less than 100, it will be better to take all of them as the sample, but if the population is more than 100 it can take 10-15 % or 20-25 % or more, depending on the researcher's ability, time, fund

and energy. Hornby (2002:745) states that sample is "part of whole", and according to Suharsimi Arikunto (2002:109), sample is part of representative of the research population.

The population of this research is the seventh grade students of SMP 10 Muhammadiyah Yoyakarta, there are 110 students divided into 3 classes. The writer chooses the sample by using purposive sampling. Margono (2005:94) states that purposive sampling technique is done based on certain features of the object research in which closely related to the known population. In other word, the sample of the study is taken based on the purpose of the study.

In this study, the writer collected the data in the form of students' answer sheets. Then the writer did scoring the students answers sheet. The aim of this study in order to find the level of a good test items criteria. So the writer chose the sample by taking the students' mark in the same distribution.

The data in this research are in the form of test item and the students' answer sheets of the first term test in the academic year of 2013/2014. The writer took the data from the school then analyzed them. There are fifty items in this test, which are divided into five parts; they are objective test, reading comprehension, arranging words, writing and completing sentence. Then the writer analyzed the data based on the literature review which is discussed before.

E. Research Variable

Variable is the conditions or characteristics that an experimenter manipulates, controls or observes. According to Keringer (1981:21), a variable is

a symbol to which a number or value is assigned. An independent variable is presumed to be the cause of dependent variable, the presumed effect. The independent variable is the attendant, while the dependent variable is the consequence.

There are two variables in this study, the English test as the instrument and the test results of it. The independent variable is the English test and the dependent variable is the test result of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta.

F. Research Instrument

The research instrument is the English test in the first term of the seventh grade student of SMP 10 Muhammadiyah Yogyakarta. Besides, the writer also uses the students' answer sheet of that test to find out the validity, reliability, difficulty index, the discrimination index and also the efficiency distractor of the test itself. First of all the writer took the score of the test result, and then analyzed them used *Item and Test Analysis (ITEMAN)*.

The instruments of the validity, reliability, difficulty index, discrimination index and efficiency distractor are as follows:

a. The instrument of validity

To find the validity of objective items, the *coefficient point biserial correlation* (Arikunto, 2001:76) is used.

$$r_p \text{ bis} = \frac{Mp - Mt}{St}\sqrt{p/q}$$

$r_p$bis is coefficient point biserial correlation

X_p is the average of the true answer's score

X_t is the average of total score

P is the proportion of the students who answer true

q is the proportion of the students who answer false

S_t is the standard deviation

The coefficient point biserial correlation from the calculation is consulted with the *r* table at significance level 5% in accordance with number of the students. If $r_pbis > r$ table, the test item is valid.

b. The instrument of reliability

The reliability of objective test will be found with the formula KR-20 as follows:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{s^2 - \sum pq}{s^2}\right)$$

r_{11} = estimated reliability for the full-length test
n = number of item
s^2 = standard deviation
Σpq = sum the product of p and q all item
p = proportion of students passing the item
q = proportion of students who failing the item (or 1-p)

(Suharsimi, 2009: 100)

Furthermore, the provision of interpretation of reliability coefficients tests (*r11*) is generally used benchmark as follows:

a. If *r11* is equal to or greater than 0.70 means the test material being tested has a high reliability or reliable.
b. If *r11* smaller than 0.70 means the test material being tested ha a low reliability index or unreliable.

(Sudijono, 2011: 209)

c. The instrument of the difficulty index

The difficulty index is found using the formula:

$$P = \frac{B}{n}$$

Where:

P = difficulty index

B = total of the subjects who answer true

n = total of all subjects

According to Arikunto (2001:210), the difficulty index is classified into:

Difficult = difficulty index (P): 0.00 until 0.3

Medium = difficulty index (P): 0.3 until 0.7

Easy = difficulty index (P): 0.7 until 1.0

d. The instrument of the discrimination index

The formula to calculate the discrimination index is as follows:

$$D = \frac{BA}{nA} - \frac{BB}{nB}$$

Where:

BA= the total of right answers for upper class

BB= the total right answers of the lower class

nA= subject of upper class

nB= subject of lower class

(Suharsimi, 2009: 214)

According to Zainal Arifin (2012: 274), the discrimination index is classified into:

D: 0, 00 - 0, 19: poor

D: 0, 20 -0, 29 : marginal

D: 0, 30 -0, 39 : good

D: 0, 40 < : very good

e. Distractor efficiency

The test material consists of two parts; those are the question and the alternative answer. The alternative answer also consists of two parts; the key answer and the distractor. A distractor is categorized as an effective distractor if the lower the student's ability the more they choose the distractor or the higher the student's ability, the less they choose the distractor.

If the proportion of the test participants who choose the wrong answer or the distractor is less than 0,025, the distractor should be revised. If the proportion is 0, 00 or there are no students who choose the distractor, the distractor should be rejected (Depdikbud : 1997). The proportion of each alternative answer can be seen from the *proportion endorsing* column in the ITEMAN analysis.

According to Depdikbud (1997), the distractor efficiency is classified into:

Good : ≥ 0,025

Revised : < 0,025

Rejected : 0,000

f. The quality of the test item

After being analyzed based on each criterion, then the test item is analyzed generally based on the validity, reliability, difficulty index, discrimination index, and the distractor efficiency to determine the quality of the test item used in the test material. The determination of the quality of the test item is based on some consideration as follows:

a. The test item categorized as a good item if the test item meet the four criteria; validity, difficulty index, discrimination index, and the distractor efficiency.

b. The test item categorized as a good enough items if the test item has three criteria from those four criteria.

c. The test item categorized as a not good item if the test item do not meet two or more criteria from those the four criteria.

G. Data Collection Technique

Data collection technique can be categorized into two groups; primary data collection and secondary data collection (Fried Schmidt, 1997: http://ers.edu/comunity data/data.html). Primary data collection technique obtains data directly from a client or target population. Secondary data collection technique obtains data from one collected sources. The writer collected the data using secondary data collection technique. Concern with the data, the researcher got the data from the object of the research; here is the school as the object of the research. The writer went to SMP 10 Muhammadiyah where the even term test of the seventh grade is taken for the sample. Then, the writer asked for the documentation of the test item and also the students' answer sheet of that test to the teacher.

H. Data Analysis Technique

After collecting the data (the test items and the students' answer sheet) the writer analyzed them. Here the writer analyzed the validity, reliability, difficulty index, discrimination index and the efficiency distractor of the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta. Validity of the objective test items is found using the biserial correlation formula. Reliability of the objective test item is calculated using the KR 20 formulas. The difficulty index of the objective test items is found using the formula of difficulty index. Discrimination index of the objective test items is

analyzed using the formula of the discrimination index. The formulas are presented in the previous chapter.

Here are the stages to analyze the data:

1. The data are in a form of multiple-choice test items of the first- term test.
2. The writer did a correction and scoring on the students' answer sheet and made it as the data to be calculated. The writer used *Item and Test Analysis* program (ITEMAN) in calculating the data. The test items are analyzed based on its validity, reliability, difficulty index, the discrimination index and the distractor efficiency.
3. After that, the writer used the statistic to make the interpretation of each item test.

CHAPTER IV

RESEARCH FINDINGS AND DISCUSSION

This chapter presents and discusses the finding of the research. This study was conducted to investigate the research problem. As stated in chapter I, the research problem is to find out the quality of the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014. In order to answer the question, this chapter is divided into three main sections. In the first section of the chapter, the description of the data is presented. The second section presents the analysis of the data, while in the last section the discussion of the data is presented.

A. Description of the data

This research was conducted to find out the quality of test item in the first term test of the second grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 which is reviewed from the validity, reliability, difficulty index, discrimination index and the efficiency distractor.

The data which were used in this research are the test items of the first term test of the second grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 that consists of 45 objective tests. There are 110 students which are divided into 3 classes who participated in this test. The data were collected using documentary method, which is in the form of the test items, and the students' answer sheet. Then the data were analyzed using *Item and Test*

Analysis Program (ITEMAN) Micro CAT version 3.00 to find out the validity, reliability, difficulty index, discrimination index, and the efficiency of the distractor.

B. Analysis of the data

1. Validity

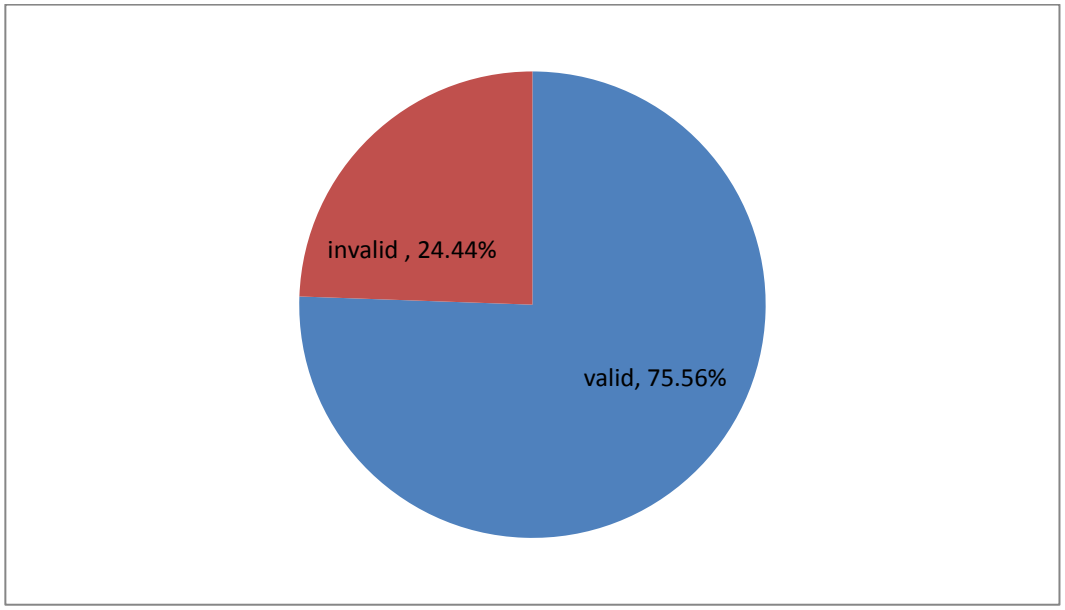
The validity of the test item was processed through the formula of correlation of biserial point ($r_p bis$). This formula calculates the items based on the students' correct answers. After scoring the students' answer sheet, the writer calculated the validity measurement and made the interpretation of this analysis. The interpretation based on the standard of *correlation coefficient* at significance level 5%. In the significance level 5%, with n= 110, it is obtained the *r* table is 0,176. If the *biserial point* $\geq$ *r* table, the test item is valid.

From the analysis of the test item of first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014, it was revealed the correlation between the item score and the total score. Then the score was consulted to the *r* table at significance level 5% and with *n* (amount of data) = 110, that is 0,176. In the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014, it was obtained that the $r_p bis$ of the test item number 2,5,12,21,22,23,25,31,36,37 and 39 is less than 0,176. It can be concluded that those test items are invalid. While the other test items have $r_p bis$ more than 0,176, so they are valid.

The distribution of the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah 10 Yogyakarta in the academic year of 2013/2014 based on the validity index is as follows:

Table 2. The distribution of the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah 10 Yogyakarta in the academic year of 2013/2014 based on the empiric validity index

No	Validity index	Test item	Total	Percentage
1.	< 0,176 (invalid)	2,5,12,21,22,23,25,31,36, 37,39	11	24,44 %
2.	≥ 0,176 (valid)	1,3,4,6,7,8,9,10,11,13,14, 15,16,17,18,19,20,24,26,27 ,28,29,30,32,33,34,35,38, 40,41,42,43,44,45	34	75,56%



Picture 1. The distribution of the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah 10 Yogyakarta in the academic year of 2013/2014 based on the empiric validity index.

2. Reliability

A test item can be considered as reliable if the $r_{11} \geq 0{,}70$, if $r_{11} < 0{,}70$ the test item considered as unreliable. The result of analysis using *ITEMAN Micro CAT Version 3.00* shows that the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 has reliability index in the amount of 0,668. The value of r_{11} which is obtained is 0,668 < 0, 70, so the test item of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 is unreliable.

3. Difficulty index

Difficulty index can be find out by proportional correct which calculated using program *ITEMAN Micro CAT Version 3*.00. The clarification that is used to

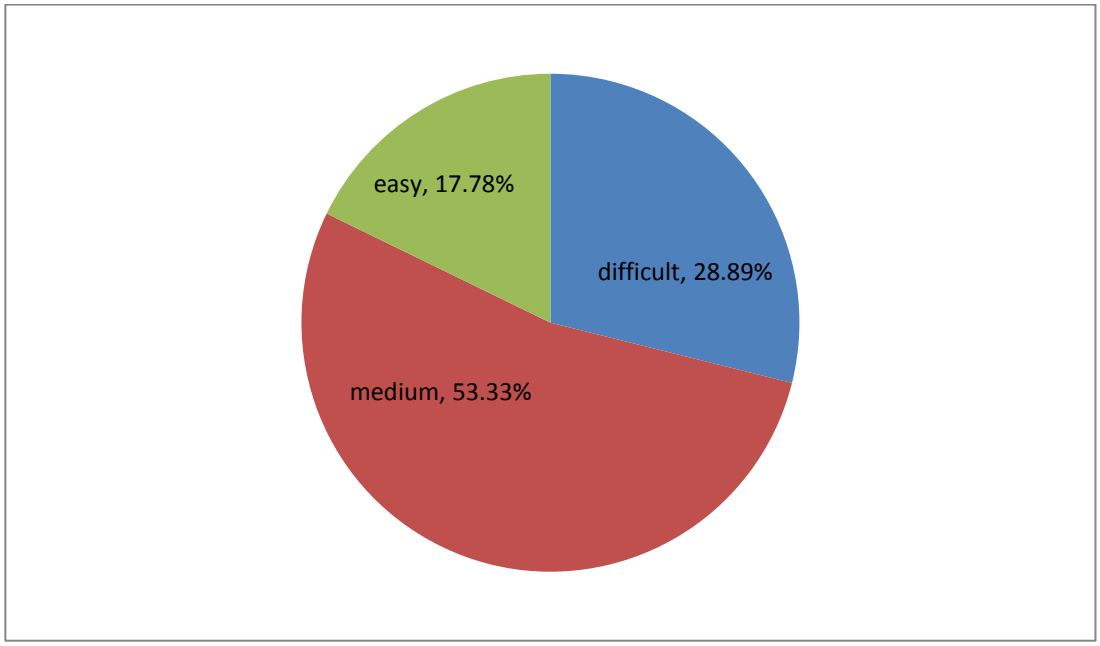
interpret the result of calculation difficulty index is; 0,00 – 0,30 be included as difficult test item; 0,31 – 0,70 be included as medium test item, and 0,71 – 1,00 be included as easy test item.

Based on the analysis using *ITEMAN Micro CAT Version 3.00* , it was find out that the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 in objective form which are categorized as difficult item come to 13 items (28,89%), categorized as medium item come to 24 items (53,33%) and categorized as easy items come to 8 items (17,78%).

The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year 2013/2014 based on the difficulty index is as follows:

Table 3. The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the difficulty index

No	Difficulty index	Test item	Total	Percentage
1.	0,00 – 0,30 (difficult)	2,5,12,13,15,16,27,28,32,36,37,39,41	13	28,89%
2.	0,31 – 0,70 (medium)	1,3,4,6,8,9,10,11,14,19,20,21,22,23,24,26,29,31,33,38,42,43,44,45	24	53,33%
3.	0,71 – 1,00 (easy)	7,17,18,25,30,34,35,40	8	17,78%



Picture 2. The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the difficulty index

4. Discrimination index

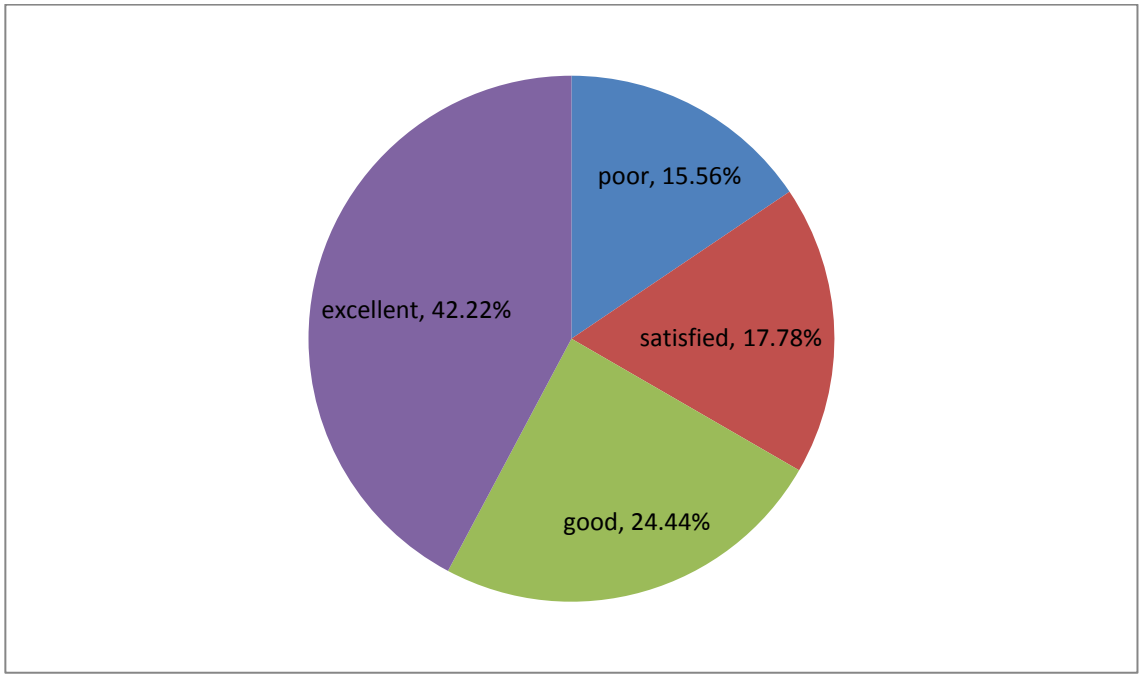
Discrimination index of test item can be finding out by look at the *biser* column in the ITEMAN analysis. The clarification that is used to interpret the result of the calculation discrimination index is: 0,00 – 0,19 categorized as poor, 0,20 – 0,29 categorized as satisfactory, 0,30 – 0,39 categorized as good, and > 0,40 categorized as excellent.

Based on the result of analysis using *ITEMAN Micro CAT Version 3.00*, it is revealed that the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta consists of 7 items (15, 56%) that have a poor discrimination index, 8 items (17, 78%) that have satisfied discrimination index, 44 items (24, 44%) that have good discrimination index, and 19 items (42,22%) that have excellent discrimination index.

The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the discrimination index is as follows:

Table 4. The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the discrimination index

No	Discrimination index	Test item	Total	Percentage
1.	0,00 – 0,19 (poor)	2,21,22,23,25,31,39	7	15,56%
2.	0,20 – 0,29 (satisfactory)	11,13,14,15,26,36,37,40	8	17,78%
3.	0,30 – 0,39 (good)	3,5,8,12,24,28,30,32,42,43,44	11	24,44%
4.	≥ 0,40 (excellent)	1,4,6,7,9,10,16,17,18,19,20,27,29,33,34,35,38,41,45	19	42,22%



Picture 3. The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the discrimination index.

5. The distractor efficiency

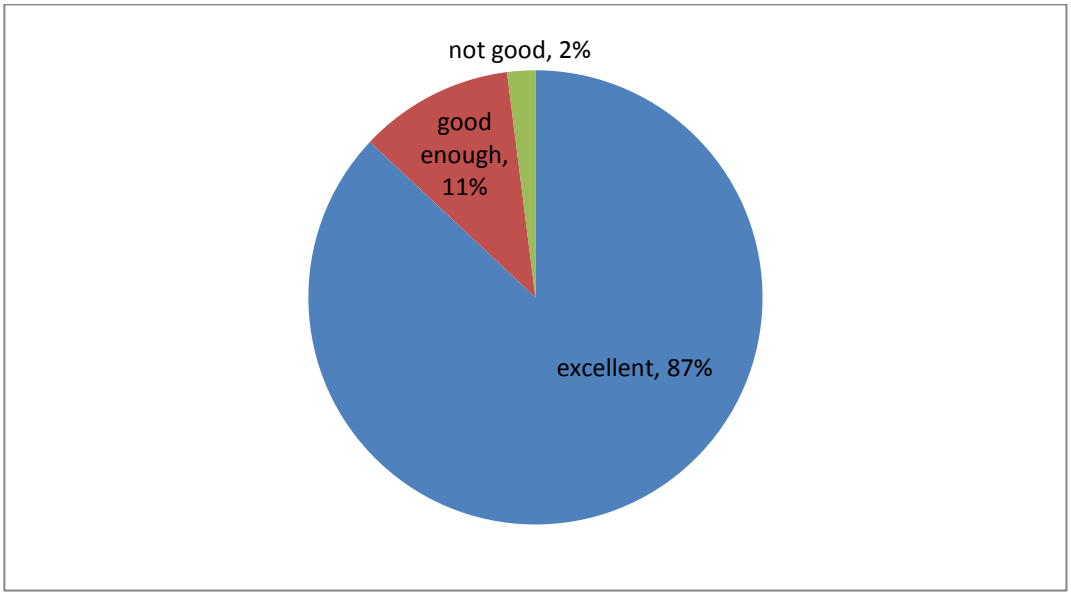
The distractor efficiency of test item can be finding out by look at the *proportion endorsing* column in the ITEMAN analysis. The clarification that is used to interpret the result of the calculation distractor efficiency is: ≥ 0, 025 categorized as good, < 0, 025 categorized as should be revised, and if the *proportion endorsing* is 0, 000 means the test item is rejected.

Based on the result of analysis using *ITEMAN MicroCAT Version 3.00*, it is revealed that the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta consists of 39 items (87%) that have an excellent distractor efficiency, 5 items (11%) that have good enough distractor efficiency and 1 item that has not good distraction (2%).

The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the distractor efficiency is as follow:

Table 5. The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the distractor efficiency

No	Effective Distractor	Test item	Total	Percentage
1.	Excellent	1,3,4,6,7,8,9,10,12,14,15,16,17,18,19,20,21,22,23,25,26,27,28,29,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45	39	87%
2.	Good enough	2,5,11,24,30	5	11%
3.	Not good	13	1	2%



Picture 4. The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the discrimination efficiency.

C. Discussion

This research is aimed to find out the quality of the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014. On the whole, the test items of the first term test of the seventh grade students of SMP 10 Yogyakarta in the academic year of 2013/2014 are poor. There are five indicators that affect the quality of test item; validity, reliability, difficulty index, discrimination index and efficiency distractor.

1. Analysis of validity

Validity refers to how well a test measures what it is purposed to measure. The validity of the test item can be calculated by calculate the validity of each test item and the correlate it with the validity of the whole test item. The number that shows validity index can be obtained by calculating correlation index between each item's score and the total score. The empiric validity is calculated using *coefficient point biserial correlation* formula, which is consulted with *r* table at significance level 5%. The students who are used as subject of this research in SMP 10 Muhammadiyah Yogyakarta are 110 students. Therefore, the value of *r* table is 0,176.

The result of the research and analysis of the test item of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the precept that $r_p bis \geq 0{,}176$ is valid, while

if r_pbis < 0,176 in invalid. The invalid test item should be repaired, while the valid test item can be used again.

Based on the analysis above, it can be concluded that the test items of the first term test of the seventh grade student of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 are considered as good test items based on the quality of the validity.

2. Analysis of reliability

Reliability is the degree to which an assessment tool produces stable and consistent result. Reliability can be measured by KR – 20 formula. If $r_{11} \geq 0{,}70$, it means the test item has a high reliability. While if $r_{11} < 0{,}70$, means the test item has a low reliability or unreliable.

The result of this research shows that the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2015 have reliability index 0,668. It means that the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 are considered as unreliable because the value of r_{11} < 0, 70 and the result will not be stable or consistent if the test is tested again in the same group.

3. Analysis of difficulty index

Difficulty index indicates the proportion of students who got the item right. A high percentage indicates an easy item/question and a low percentage indicates a difficult item. A test can be considered as a good test if it not too easy or not too difficult.

Based on the research, in the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 there are 13 items (28,89%) that categorized as difficult items, 24 items (53,33%) that medium difficult items, and 8 items (17,78%) that categorized as easy items.

From the analysis above, it can be concluded that the test items of the English first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 are categorized as good test items because most of the test items are medium difficult (53,33%). Those test items which have medium difficulty can be used again in the next evaluation. While test items which difficult or easy can be omitted and do not used again in the next evaluation, or can be revised and traced what factor which affect why those item can not be answered by students and why those items are too easy to answer by students.

4. Analysis of discrimination index

Discrimination index is the difference between the proportion of the top scorers who got an item correct and the proportion of the bottom scorers who got the item right. Based on the research finding, in the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 there are 7 items (15,56%) which have poor discrimination index, 8 items (17,78%) which have satisfied discrimination index, 11 items (24,44%) which are categorized as good discrimination index and 19 items (42,22%) that have excellent discrimination index.

From the analysis above, it can be concluded that the test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 are good, because 42,22% of the items have excellent discrimination index.

5. Analysis of distractor efficiency

A distractor is categorized as an effective distractor if the lower the student's ability the more they choose the distractor or the higher the student's ability, the less they choose the distractor.

If the proportion of the test participants who choose the wrong answer or the distractor is less than 0,025, the distractor should be revised. If the proportion is 0, 00 or there are no students who choose the distractor, the distractor should be rejected (Depdikbud : 1997). The proportion of each alternative answer can be seen from the *proportion endorsing* column in the *ITEMAN* analysis.

From the analysis above, it can be concluded that the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 is good, because 87% of the items have excellent distractor efficiency.

6. Analysis of the test item based on the validity, reliability, difficulty index, discrimination index, and the distractor efficiency

After being analyzed based on each criterion, then the test item is analyzed generally based on the validity, reliability, difficulty index, discrimination index, and the distractor efficiency to determine the quality of the test item used in the

test material. The determination of the quality of the test item is based on some consideration as follows:

a. The test item categorized as a good item if the test item meet the four criteria; validity, difficulty index, discrimination index, and the distractor efficiency.

b. The test item categorized as a good enough items if the test item has three criteria from those four criteria.

c. The test item categorized as a not good item if the test item do not meet two or more criteria from those the four criteria.

The result of the general analysis of the test item of the first term test of the seventh grade students of SMP Muhammadiyah 10 Yogyakarta in the academic year of 2013/2014 based on the validity, reliability, difficulty index, discrimination index, and the distractor efficiency is as follows:

Table 6. The distribution of test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 based on the validity, reliability, difficulty index, discrimination index, and the distractor efficiency.

No	Quality	Item	Total	Percentage
1.	Good	1,3,4,6,9,10,19,20,29,33,38 ,45,	12	27%
2.	Good enough	7,8,14,16,17,18,26,27,34,35,4 1,42,43,44	14	31%
3.	Poor	2,5,11,12,13,15,21,22,23,24,2 5,28,30,31,32,36,37,39,40	19	42%

Based on the table above, it can be revealed that the test item of the first term test of the seventh grade students of SMP Muhammadiyah 10 Yogyakarta in the academic year of 2013/2014 has a poor quality because 42% of the test items are categorized as poor. The test items that have a good quality can be used again in the next examination, the items that have a good enough quality can be revised and be used again, while the items that have a poor quality should be omitted and do not use again in the next examination. For example, the test item number 8 and 9, they have good and good enough quality based on the some consideration above. While based on the empirical validity analysis, they are valid. But based on the construct validity, they are not valid, because there are grammatical errors in the question sentences. It means that the test items number 8 and 9 can be used again in the next evaluation but they need to be revised. While for test item number 11, it has a poor quality based on quantitative analysis because it has a poor efficiency distractor. It also invalid viewed from the construct validity, because it is grammatically error. So item number 11 should be omitted and do not use again in the next evaluation.

The failed of the test items is caused by there are some criteria that are not fulfilled. The following is the caused of the failed test item:

Table 7. The caused of the failed test items of the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014.

The caused	Test item	Total	Percentage
Validity	2,5,12,21,22,23,25,31,36,	11	24,44 %

	37,39		
Difficulty index	2,5,12,13,15,16,27,28,32,36,37 ,39,41	13	28,89%
Discrimination index	2,21,22,23,25,31,39	7	15,56%
Distractor efficiency	2,5,11,24,30,13	6	13%

Based on the table above, can be concluded that the main cause of the failed test is the difficulty index that means the test item is too difficult or too easy. A good test item should have a medium difficulty. The second cause is the validity that means that the test item does not have alignment with the total score. Then the next cause is the discrimination index, it means that the test item of the first term test of the seventh grade students of SMP Muhammadiyah 10 Yogyakarta can not differentiate the students who have high ability and the students who have low ability. And the last cause is the distractor efficiency, it means that test item have the distractor which can not function well. A distractor can be categorized as a good distractor if the distractor can "lure" a sufficient number of test-takers especially lower ability ones.

The test item that has good enough and poor quality can be used in the next evaluation with some revision based on the cause of the failure or can be omitted. While the test item that has a good quality can be used again in the next examination without revision.

CHAPTER V

CONCLUSION AND SUGGESTION

In this chapter, the writer presents the conclusions of the study and some related suggestions.

A. CONCLUSION

This research is conducted to reveal the quality of the test item. The item with good quality must be kept and can be used for the next test. On the contrary, the items with poor quality must be taken away and replaced by good items. The characteristics of a good test are having high measurement of validity, reliability, difficulty index, discrimination index, and the efficiency distractor. Best test items can be identified through a test item analysis. The factors that can be identified through an analysis of test items are validity, reliability, discrimination index, the difficulty index and the efficiency distractor.

The item analysis has some advantages. First, it helps the test developers to identify the good or poor test items. By analyzing the test items, the test makers got the information to complete the test items for the next test. Secondly, test developers get the description about the test that they developed. Analyzing tests items can be done to objectives test items. There is an appropriate formula and the way to analyze the test item based on its criteria.

The conclusions presented below are drawn in accordance with the problems of the study.

1. The results of the study show that most of the test items of the first term test item for the seventh grade students of SMP 10 Muhammadiyah Yogyakarta are valid. There are 75, 56% of the test items which have $r_p bis \geq 0{,}176$. It means that the validity measurement is satisfactory.
2. The result of the study show the first -term test of the seventh grade students of SMP Muhammadiyah 10 Yogyakarta has *alpha coefficient* 0,668. It means the status of the test is unreliable.
3. The results of the study shows that the test items of the English first term test of the seventh grade students of SMP Muhammadiyah 10 Yogyakarta that have difficult items are 13 items (28,89%). The medium items are 24 items (53,33%) and the easy items are 8 items (17,78). It means that this test is not so difficult and not so easy either. Most of the items have normal difficulty index. So based on the test scores, the writer concludes that this test items having good difficulty index.
4. The first term-test items of the first - term test item for the seventh grade students of SMP 10 Muhammadiyah Yogyakarta consist of 45 items. Those items consist of 42, 22% excellent, 24, 44% good, 17, 78% satisfactory and 15,56% that includes as poor discrimination indexes category. Based on the analysis results the writer concludes that the test items having good discrimination index.

5. The finding of this research shows that 87% of the items have excellent distractor efficiency and 13% of the test items have good distractor efficiency. It can be concluded that the first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta in the academic year of 2013/2014 is good, because most of the distractor work well.

The first term test of the seventh grade students of SMP 10 Muhammadiyah Yogyakarta are developed by team. They have some strength and weaknesses. The strengths are as follows; the test is not so difficult and not so easy either. Most of the items have normal difficulty index. The weaknesses of the test item are as follows; the objective tests of these items are unreliable. The *alpha* is 0,668 < 0, 70. The result of the test item analysis shows that the quality of the test items of the English first term test of the seventh grade students of SMP Muhammadiyah 10 Yogyakarta is poor. It means that those test items need to be improved.

B. SUGGESTION

In relation to the evidence found in test items analysis, the writer put forward some suggestions as follows:

1. To the English teacher

   The writer considers that the teacher should pay attention to the previous test and conduct the test item analysis first. They should have

the ability to analyze the test items. The result of item analysis can be used to select items of desired difficulty that best discriminate between high and low achieving students. However the results of an item analysis can be useful in identifying faulty items and can provide information about student misconceptions and topics that need additional work. Then, they must be able to create and delivered the material based on the curriculum. So, later on, when the students face the term test they will not find difficulties.

2. To the Test Developers

   The writer suggests that in making the test, especially for the large scopes, they should not forget to analyze the test they develop. By doing so, the test they developed will become a standardized test.

3. To the Educational and teacher training education program

   The writer suggests that the department runs an instructional program about the system of test development. Such instruction should include the running of test-item analysis.

4. To other researcher

   The writer also suggests other researchers to use this study as a reference in order to be able to make a more comprehensive research.

References

.

Arikunto,Suharsimi. (2009) . *Dasar-dasar Evaluasi Pendidikan*. Jakarta: BUMI AKSARA

Anderson, Scarvia B. (1976). *Encyclopedia of Educational Evaluation*. London: Jossey Bass Publisher.

Brown, H. Douglas. (2004). *Language Assessment Principles and Classroom Practices*. USA: Pearson Education,Inc.

Brown, H. Douglas. (2005). *Testing in Language Programs.* New York.McGraw-Hill Companies.

Hadi, Sutrisno. (2000). *Seri Progam Statistik*. Yogyakarta: Universitas Gajah Mada.

Hopkins,C.D.and Richard, L.A. (1964). *Classroom Measurement Evaluation*. New Delhi: Indiana State University Press.

Jaeger,Richard.M. (1983). *Statistics (A Spectator Sport)*. London: SAGE PUBLICATIONS.

Margono. (2005). *Metodologi Penelitian Pendidikan (fifth edition)*. Jakarta : Rineka Cipta.

Nurgiyantoro, B. (1987). *Penilaian dan Pengajaran Bahasa dan Sastra.* Yogyakarta: BPFE.

Departemen Pendidikan Dan Kebudayaan. (1997). *Manual Item and Test Analysis (Iteman)*. Badan Penelitian dan Pengembangan Pendidikan dan Kebudayaan: Pusat Penelitian dan Pengembangan Sistem Pengujian.

Purwanto, Ngali. (1991). *Prinsip- Prinsip dan Teknik Evaluas*i *Pengajaran*. Bandung: Pemuda Rosdakarya.

Sukmadinata, N.S. (2008). *Metode Penelitian Pendidikan (edisi ke 4).* Bandung. PT REMAJA ROSDAKARYA

Trip Frach. *Test and Item Analysis*. Retrieved (Sunday, 17th Oct 2010) from http://www.inovasi.lipi.go.id/elearning/help.php?module=quiz&file=itemanalysis.html

APPENDICES

APPENDIX A

PIMPINAN WILAYAH MUHAMMADIYAH MAJELIS DIKDASMEN DIY
BKS SMP MUHAMMADIYAH DAERAH ISTIMEWA YOGYAKARTA
ULANGAN AKHIR SEMESTER GASAL BERSAMA
TAHUN AJARAN 2013 / 2014

MATA PELAJARAN	:	BAHASA INGGRIS
KELAS	:	VII (TUJUH)
HARI / TANGGAL	:	Rabu, 11 Desember 2013
PUKUL	:	07.30 - 09.30 WIB

PETUNJUK UMUM :

1. Berdoalah sebelum dan sesudah mengerjakan soal.
2. Perhatikanlah dan ikuti petunjuk pengisian pada lembar jawaban yang disediakan.
3. Periksa dan bacalah setiap soal ulangan dengan seksama sebelum menjawab.
4. Laporkan kepada pengawas kalau terdapat tulisan yang kurang jelas, rusak atau jumlah soal kurang.
5. Mintalah kertas buram kepada pengawas apabila diperlukan.
6. Kerjakan pada lembar jawaban yang disediakan dengan menggunakan ballpoint atau tinta.
7. Apabila anda ingin memperbaiki/mengganti jawaban cukup dengan memberi tanda samadengan (=) pada huruf yang akan diganti.
8. Periksalah seluruh pekerjaan anda sebelum diserahkan kepada pengawas.

SELAMAT MENGERJAKAN

PETUNJUK KHUSUS

A. PILIHAN GANDA

I. For questions number 1 s/d 45 choose the correct answer by crossing A, B, C or D on the answer sheet.

The following text is for questions 1 - 3

Hello everyone. I'd like to introduce myself. My name is Ibrahim Maulana Hafid. You can call me Ibra. I am thirteen years old. I study at SMP Muhammadiyah Cinta Tanah Air in the seventh grade. I live at Jalan KHA Dahlan number 12, Yogyakarta. During my spare time I like helping my dad clean our house and do gardening. My hobby is playing badminton and reading.
That's my short introduction. Thank you.

1. What is the purpose of the text?
   A. To describe the speaker's family.
   B. To introduce the speaker's himself.
   C. To describe the speaker's address.
   D. To inform about the speaker's hobby.

2. What does Ibra usually do during his spare time?
   A. Playing badminton.
   B. Helping my dad.
   C. clean his house.
   D. Reading.

3. "During my <u>spare time</u> I like helping my mom ..."
   The synonym of the underlined word is ....
   A. time table
   B. busy time
   C. free time
   D. on time

The following text is for questions 4 – 6

Attention, please. Tomorrow is May 2, 2013. We will have a flag hoisting ceremony to commemorate National Education Day. Don't forget to wear your caps and white and blue uniforms. The ceremony will begin at seven sharp. Please arrive at school at least 5 minutes earlier. Thank you.

4. What is the announcement about?
   A. A flag hoisting ceremony.
   B. White and blue uniforms.
   C. Arriving at school earlier.
   D. National Education Day.

5. What is the date today?
   A. May 4, 2013.
   B. May 3, 2013.
   C. May 2, 2013
   D. May 1, 2013

6. What time should the students arrive at school?
   A. At 6.45 a. m.
   B. At 6.55 a. m.
   C. at 7.05 a. m.
   D. At 7.10 a. m.

The following schedule is for questions 7 – 10

VII A Lesson Schedule

Period	Time	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1	07.00-07.50	Flag Ceremony	Sports	English	Mathematics	Social studies	IT
2	07.50-08.30	Indonesian	Sports	English	Mathematics	Social studies	IT
3	08.30-09.10	Indonesian	Mathematics	Social studies	Science	English	Mathematics
	09.10-09.40	BREAK					
4	09.40-10.20	Science	Mathematics	Social studies	Science	English	Mathematics
5	10.20-11.00	Science	Religion	Culture and Art	Indonesian	Religion	Councelling
6	11.00-11.40	Civics	Javanese	Culture and Art	Indonesian	-	Councelling
7	11.40-12.20	Civics	Javanese	-	-	-	-

7. What is the schedule about?
   A. Kinds of lesson.
   B. The writer's lesson.
   C. VII A lesson schedule.
   D. The day and the time of lessons.

8. What lesson do VII A have on Friday in the third period?
   A. Culture and Arts.
   B. Social Science.
   C. Mathematics.
   D. English.

9. In which period do VII A have Indonesian on Thursday?
   A. The second and third.
   B. The third and fourth.
   C. The fourth and fifth.
   D. The fifth and sixth.

10. What time does Mathematics finish on Thursday?
    A. At half past eight.
    B. At twenty minutes past ten.
    C. At eleven.
    D. At Eleven forty.

The following text is for questions 11 – 14

Yesterday I went to my friend's house. I meet a little boy there. I ask Hafid about him. Hafid says:

"Let me tell you about my brother. He is Sultan Annaja. But you can call him Sultan. He is only seven years old, five years younger than me. He is in the first grade now and is very happy about going to school. He is very cute with his chubby cheeks. He is also a clever student. He always gets good marks on his test. I really love him".

11. What is the writer's fiend name?
    A. Hafid.
    B. Sultan.
    C. Annaja.
    D. Sultan Annaja.

12. Who is Sultan Annaja? He is the writer's ....
    A. friend
    B. brother
    C. teacher
    D. friend brother

13. How old is Hafid? Hafid is ... years old.
    A. six
    B. seven
    C. twelve
    D. thirteen

14. "He is also a clever student."
    The underlined word has the same meaning to ....
    A. diligent
    B. careful
    C. smart
    D. active

The following dialog is for questions no 15 – 17

Adnin : Do you have brother or sister, Aulia?
Aulia : Yes. I have one elder sister, Namira, one younger sister, Marwa and one younger brother, Azka.
What about you?
Adnin : I have only one elder sister, Habiba.

15. Aulia is the ... child in his family.
    A. first
    B. second
    C. third
    D. fourth

16. How many siblings does Aulia have?
    A. One.
    B. Two.
    C. Three.
    D. Four.

17. Habiba is ...sister
    A. Adnin's
    B. Aulia's
    C. Marwa's
    D. Namira's

The following information is for questions no 18 – 20.

Name	:	Alkalifi Sultan Annaja
Age	:	14
Parents	:	Ommand Cakra Wijaya & Daisy Rizqi Putri
Siblings	:	Aurelia S Adnin (elder sister)
		Alisha S Annalani (elder sister)
		Adyana Shasmika A (younger sister)
School	:	SMP Pancasila Sakti
Adress	:	Jalan Wahid Hasyim no 17 Yogyakarta
Hobby	:	Philately

18. Who is Alkalifi's mother?
    A. Aurelia
    B. Adyana
    C. Alisha
    D. Daisy

19. Alkalifi is the only ... of Ommand and Daisy
    A. daughter
    B. brother
    C. sister
    D. son

20. Alkalifi's hobby is philately.
    The underlined word means ....
    A. collecting old items
    B. correspondensing
    C. collecting stamps
    D. writing letters

The following schedule is for questions no 21 – 23

Safa's Daily Schedule

Time	Activity
04.30 a.m.	Get up and make a bed.
04.30 a.m. – 05.00 a.m.	Take a bath, get dressed and do morning prayer.
05.15 a.m.- 05.45 a.m.	Have breakfast
06.00 a.m.	Leave for school
07.00 – 12.30	Study at school

21. What time does Safa usually get up?
    A. A quarter to four.
    B. A quarter past four.
    C. A half to five.
    D. A half past five.

22. How long does Safa have for study at school?
    A. Five hours.
    B. Five and a half hours.
    C. six hours.
    D. six and a half hours.

23. From the schedule we know that Safa ....
    A. needs twenty minutes to have breakfast
    B. does morning prayer after take a bath
    C. always gets up late every morning
    D. goes to school at seven o'clock

The following text is for questions no 24– 26

To : Aflah
Don't forget that we'll be going jogging with Haris tomorrow morning. Remind Bintang and Izza about it. Asked them to gather at my house at 5 a.m. Thank you. C U.

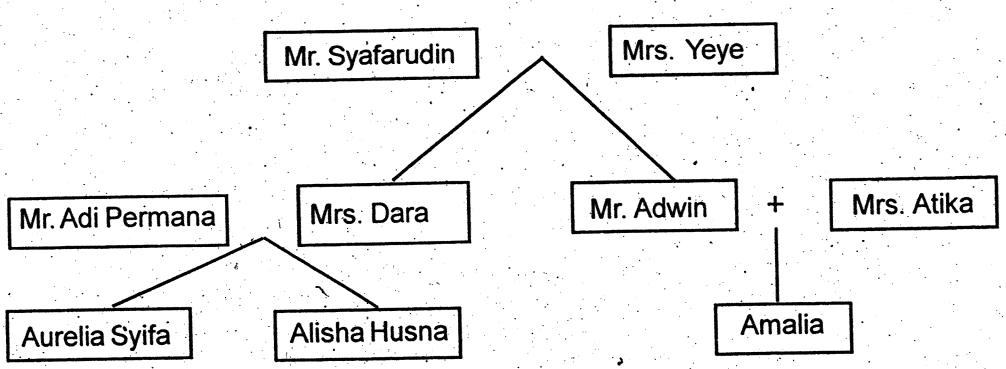
Your best friend.
Satya.

24. Who is the message from?
    A. Satya..
    B. Haris..
    C. Aflah
    D. Bintang

25. How many boys will go jogging?
    A. Three.
    B. Four.
    C. Five.
    D. Six.

26. "Asked <u>them</u> to gather at my house at 5 a. m."
    What does the underlined refer to?
    A. Bintang and Izza.
    B. Aflah and Satya.
    C. Aflah and Haris.
    D. Satya and Izza.

The following family tree is for questions no 27– 29



27. What is the relationship between Mr. Syafarudin and Mr. Adwin?
    Mr. Adwin is Mr. Syafarudin's ....
    A. grandpa.
    B. cousin.
    C. father.
    D. son.

28. Husna is Mrs.Atika's ....
    A. nephew
    B. cousin
    C. sister
    D. niece

29. From the family tree we know that Mrs. Yeye has ... grandchildren.
    A. two
    B. three
    C. four
    D. five

The following notice is for questions no 30– 32

WELCOME TO OUR SCHOOL
VISITORS MUST REGISTER AT THE MAIN OFFICE

30. Where would you probably find the notice?
    A. At home.
    B. In a shop.
    C. At school.
    D. In an office.

31. Who is the notice addressed to?
    A. Visitors.
    B. School staffs.
    C. Officers.
    D. Visitors.

32. "<u>Visitors</u> must register at the main office."
    The underlined word has the similar meaning to ....
    A. guests
    B. teachers
    C. students
    D. members

The following text is for questions no 33 – 35

Amanda's shopping list.

- Ribbons ..... yellow (20 m), blue (15 m), Red (25 m), white (10 m)
- Balloons ..... one pack
- Paper hats 48
- Paper plates 48
- Straws ..... one pack
- Tart cake ..... one
- Biscuits ..... 3 tins
- Candies ..... 2 packs
- Cocopandan syrup ..... 3 bottles

33. How many kinds of ribbons will Amanda buy?
    A. one
    B. two
    C. three
    D. four

34. Amanda needs ... paper hats.
    A. fifteen
    B. twenty
    C. twenty five
    D. forty eight

35. From the text we know that Amanda will hold ...
    A. a school exam
    B. a birthday party
    C. a beauty contest
    D. a speech contest

The following text is for questions no 36 - 38

Dear Farhan,

Tania and I want to go to Surabaya to visit uncle Sulaiman on Saturday. Would you like to join us? Please call me as soon as possible. Thanks.

Love,
Nanda

36. Who is the message written for?
    A. Sulaiman
    B. Tania.
    C. Farhan.
    D. Nanda

37. What should Farhan do?
    A. Go to Surabaya.
    B. visit uncle Sulaiman
    C. Join Tania and Nanda
    D. call Nanda immediatelly

38. "Please call <u>me</u> as soon as possible."
    The underlined word refers to ....
    A. Tania
    B. Nanda
    C. Farhan
    D. Sulaiman

For questions 39 – 41, choose the correct words to complete the paragraph.

Fatah, there are two (39) ... on your desk. One dictionary (40) ... thick, but the other is (41) .... So, which one is yours? Please reply soon. Thanks, Zuleha.

39. A. dictionary
    B. dictionarys
    C. dictionaries
    D. dictionaryes

40. A. is
    B. be
    C. am
    D. are

41. A. long
    B. small
    C. short
    D. thin

42. Rearrange these following words to make a good sentence.
    always – up – every – Malika – gets – morning – early
    1 2 3 4 | 5 6 7

    A. 4 – 3 – 7 – 2 – 5 – 1 – 6
    B. 4 – 7 – 2 – 5 – 1 – 3 – 6
    C. 4 – 7 – 1 – 5 – 3 – 2 – 6
    D. 4 – 1 – 5 – 2 – 7 – 3 – 6

43. Rearrange these following words to make a good sentence.
    The – keep – place – clean – tidy – please – and
    1 2 3 4 5 6 7

    A. 6 – 1 – 3 – 2 – 4 – 7 – 5
    B. 6 – 2 – 1 – 3 – 4 – 7 – 5
    C. 2 – 1 – 3 – 6 – 5 – 7 – 4
    D. 2 – 3 – 1 – 4 – 5 – 7 – 4

44. Rearrange these following sentences to make a meaningful message.
    1. I will pick you up at five.
    2. Thank you.
    3. Anna, this is Rosita.
    4. Please be ready.
    5. There will be a study group tomorrow.

    A. 3 – 5 – 1 – 4 – 2
    B. 3 – 1 – 2 – 5 – 4
    C. 5 – 1 – 4 – 2 – 3
    D. 5 – 2 – 3 – 1 – 4

45. Rearrange these following sentences into a meaningful announcement.
    1. Thank you.
    2. Good afternoon, all students.
    3. Each class should present one team.
    4. To celebrate National Sports Day, there will be an interclass football competition .
    5. For the schedule, please check the announcement board.

    A. 5 – 3 – 1 – 4 – 2
    B. 5 – 3 – 4 – 1 – 2
    C. 2 – 4 – 3 – 5 – 1
    D. 2 – 4 – 1 – 5 – 3

ANSWER KEY

1. B
2. C
3. C
4. A
5. D
6. B
7. C
8. D
9. D
10. A
11. A
12. D
13. C
14. C
15. B
16. C
17. A
18. D
19. D
20. C
21. C
22. B
23. B
24. A
25. C
26. A
27. D
28. D
29. B
30. C
31. A
32. A
33. D
34. D
35. B
36. C
37. D
38. B
39. C
40. A
41. D
42. D
43. B
44. A
45. C

APPENDIX B

```
3  1  Scores for examinees from file TEST1.TXT
001  25.00
002  20.00
003  18.00
004  19.00
005  17.00
006  21.00
007  15.00
008  12.00
009  17.00
010  15.00
011  17.00
012  21.00
013  15.00
014  10.00
015  28.00
016  16.00
017  16.00
018  16.00
019  13.00
020  16.00
021  21.00
022  23.00
023  24.00
024  18.00
025  19.00
026  17.00
027  22.00
028   6.00
029  14.00
030  24.00
031  17.00
032  24.00
033  17.00
034  12.00
035  19.00
036  14.00
037  21.00
038  17.00
039  24.00
040  19.00
```

041 14.00
042 32.00
043 20.00
044 16.00
045 27.00
046 30.00
047 16.00
048 26.00
049 20.00
050 19.00
051 25.00
052 25.00
053 21.00
054 20.00
055 19.00
056 20.00
057 37.00
058 31.00
059 22.00
060 14.00
061 23.00
062 25.00
063 23.00
064 22.00
065 24.00
066 21.00
067 27.00
068 12.00
069 15.00
070 26.00
071 20.00
072 16.00
073 28.00
074 23.00
075 18.00
076 18.00
077 20.00
078 33.00
079 23.00
080 18.00
081 25.00

082 21.00
083 22.00
084 23.00
085 20.00
086 19.00
087 11.00
088 19.00
089 25.00
090 19.00
091 18.00
092 17.00
093 23.00
094 14.00
095 20.00
096 23.00
097 18.00
098 17.00
099 30.00
100 25.00
101 22.00
102 23.00
103 24.00
104 22.00
105 20.00
106 22.00
107 25.00
108 20.00
109 22.00
110 23.00

APPENDIX C

MicroCAT (tm) Testing System

Item and Test Analysis Program -- ITEMAN (tm) Version 3.00

Item analysis for data from file TEST1.TXT Page 1

		Item Statistics			Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
1	0-1	0.673	0.639	0.492	A	0.045	-0.490	-0.225	
					B	0.673	0.639	0.492	*
					C	0.155	-0.475	-0.312	
					D	0.127	-0.340	-0.213	
					Other	0.000	-9.000	-9.000	
2	0-2	0.164	0.070	0.046	A	0.245	-0.339	-0.248	
					B	0.573	0.273	0.217	?
		CHECK THE KEY			C	0.164	0.070	0.046	*
		C was specified, B works better			D	0.018	-0.395	-0.132	
					Other	0.000	-9.000	-9.000	
3	0-3	0.491	0.316	0.252	A	0.118	-0.392	-0.240	
					B	0.145	-0.094	-0.061	
					C	0.491	0.316	0.252	*
					D	0.245	-0.085	-0.062	
					Other	0.000	-9.000	-9.000	

4	0-4	0.455	0.466	0.371	A	0.455	0.466	0.371	*
					B	0.100	-0.434	-0.254	
					C	0.082	-0.303	-0.167	
					D	0.364	-0.167	-0.130	
					Other	0.000	-9.000	-9.000	
5	0-5	0.045	0.307	0.141	A	0.009	-0.397	-0.103	
					B	0.027	-0.205	-0.079	
					C	0.918	-0.045	-0.025	
					D	0.045	0.307	0.141	*
					Other	0.000	-9.000	-9.000	
6	0-6	0.355	0.462	0.360	A	0.118	-0.126	-0.077	
					B	0.355	0.462	0.360	*
					C	0.491	-0.340	-0.271	
					D	0.036	-0.147	-0.063	
					Other	0.000	-9.000	-9.000	
7	0-7	0.727	0.584	0.436	A	0.045	-0.338	-0.155	
					B	0.055	-0.446	-0.217	
					C	0.727	0.584	0.436	*
					D	0.173	-0.440	-0.297	
					Other	0.000	-9.000	-9.000	

8	0-8	0.691	0.326	0.249	A	0.027	-0.118	-0.046
					B	0.209	-0.255	-0.180
					C	0.073	-0.248	-0.132
					D	0.691	0.326	0.249 *
					Other	0.000	-9.000	-9.000
9	0-9	0.573	0.408	0.323	A	0.109	-0.013	-0.008
					B	0.173	-0.376	-0.254
					C	0.145	-0.269	-0.174
					D	0.573	0.408	0.323 *
					Other	0.000	-9.000	-9.000
10	0-10	0.673	0.433	0.333	A	0.673	0.433	0.333 *
					B	0.136	-0.195	-0.124
					C	0.100	-0.527	-0.309
					D	0.091	-0.129	-0.073
					Other	0.000	-9.000	-9.000
11	0-11	0.455	0.242	0.193	A	0.455	0.242	0.193 *
					B	0.100	-0.476	-0.278
					C	0.009	-0.249	-0.064
					D	0.436	-0.016	-0.013
					Other	0.000	-9.000	-9.000

12	0-12	0.055	0.325	0.158	A	0.218	-0.208	-0.148
					B	0.700	0.084	0.063
					C	0.027	-0.060	-0.023
					D	0.055	0.325	0.158 *
					Other	0.000	-9.000	-9.000
13	0-13	0.118	0.296	0.181	A	0.018	-0.558	-0.186
					B	0.855	-0.248	-0.161
		CHECK THE KEY			C	0.118	0.296	0.181 *
		C was specified, D works better			D	0.009	0.935	0.242?
					Other	0.000	-9.000	-9.000
14	0-14	0.582	0.273	0.216	A	0.173	-0.227	-0.153
					B	0.118	-0.190	-0.117
					C	0.582	0.273	0.216 *
					D	0.127	-0.053	-0.033
					Other	0.000	-9.000	-9.000
15	0-15	0.191	0.263	0.182	A	0.255	-0.222	-0.163
					B	0.191	0.263	0.182 *
					C	0.282	0.015	0.011
					D	0.273	-0.016	-0.012
					Other	0.000	-9.000	-9.000

16	0-16	0.245	0.491	0.359	A	0.445	-0.408	-0.324
					B	0.191	-0.084	-0.058
					C	0.245	0.491	0.359 *
					D	0.118	0.149	0.091
					Other	0.000	-9.000	-9.000
17	0-17	0.709	0.411	0.310	A	0.709	0.411	0.310 *
					B	0.164	-0.254	-0.169
					C	0.064	-0.517	-0.264
					D	0.064	-0.110	-0.056
					Other	0.000	-9.000	-9.000
18	0-18	0.764	0.603	0.438	A	0.136	-0.452	-0.288
					B	0.055	-0.364	-0.177
					C	0.045	-0.490	-0.225
					D	0.764	0.603	0.438 *
					Other	0.000	-9.000	-9.000
19	0-19	0.382	0.484	0.380	A	0.164	-0.078	-0.052
					B	0.245	-0.253	-0.185
					C	0.209	-0.299	-0.211
					D	0.382	0.484	0.380 *
					Other	0.000	-9.000	-9.000

20	0-20	0.336	0.465	0.359	A	0.255	-0.256	-0.188
					B	0.127	-0.105	-0.066
					C	0.336	0.465	0.359 *
					D	0.282	-0.195	-0.146
					Other	0.000	-9.000	-9.000
21	0-21	0.355	-0.128	-0.099	A	0.091	-0.129	-0.073
					B	0.309	0.003	0.002
		CHECK THE KEY			C	0.355	-0.128	-0.099 *
		C was specified, D works better			D	0.245	0.214	0.157?
					Other	0.000	-9.000	-9.000
22	0-22	0.309	0.101	0.077	A	0.082	-0.387	-0.214
					B	0.309	0.101	0.077 *
		CHECK THE KEY			C	0.245	-0.195	-0.143
		B was specified, D works better			D	0.364	0.225	0.175?
					Other	0.000	-9.000	-9.000
23	0-23	0.364	0.191	0.149	A	0.109	-0.033	-0.020
					B	0.364	0.191	0.149 *
					C	0.300	-0.010	-0.008
					D	0.227	-0.205	-0.148
					Other	0.000	-9.000	-9.000

24	0-24	0.664	0.311	0.240	A	0.664	0.311	0.240 *
					B	0.091	-0.229	-0.130
					C	0.227	-0.235	-0.169
					D	0.018	-0.111	-0.037
					Other	0.000	-9.000	-9.000
25	0-25	0.782	-0.359	-0.256	A	0.082	0.285	0.157
					B	0.109	0.521	0.312?
		CHECK THE KEY			C	0.782	-0.359	-0.256 *
		C was specified, B works better			D	0.027	-0.551	-0.213
					Other	0.000	-9.000	-9.000
26	0-26	0.664	0.246	0.190	A	0.664	0.246	0.190 *
					B	0.236	-0.091	-0.066
					C	0.045	-0.111	-0.051
					D	0.055	-0.463	-0.225
					Other	0.000	-9.000	-9.000
27	0-27	0.282	0.445	0.334	A	0.136	-0.195	-0.124
					B	0.136	-0.195	-0.124
					C	0.445	-0.165	-0.131
					D	0.282	0.445	0.334 *
					Other	0.000	-9.000	-9.000

28	0-28	0.209	0.370	0.262	A	0.282	-0.007	-0.005
					B	0.209	-0.002	-0.002
					C	0.300	-0.298	-0.226
					D	0.209	0.370	0.262 *
					Other	0.000	-9.000	-9.000
29	0-29	0.491	0.429	0.343	A	0.318	-0.100	-0.077
					B	0.491	0.429	0.343 *
					C	0.145	-0.467	-0.303
					D	0.045	-0.300	-0.138
					Other	0.000	-9.000	-9.000
30	0-30	0.782	0.343	0.245	A	0.036	-0.375	-0.160
					B	0.018	-0.355	-0.119
					C	0.782	0.343	0.245 *
					D	0.164	-0.225	-0.150
					Other	0.000	-9.000	-9.000
31	0-31	0.418	0.196	0.156	A	0.418	0.196	0.156 *
					B	0.291	0.023	0.017
					C	0.218	-0.195	-0.139
					D	0.073	-0.196	-0.104
					Other	0.000	-9.000	-9.000

32	0-32	0.136	0.328	0.209	A	0.136	0.328	0.209	*
					B	0.200	-0.123	-0.086	
					C	0.536	-0.093	-0.074	
					D	0.127	-0.001	-0.000	
					Other	0.000	-9.000	-9.000	
33	0-33	0.673	0.453	0.348	A	0.118	-0.328	-0.201	
					B	0.127	-0.236	-0.148	
					C	0.082	-0.327	-0.181	
					D	0.673	0.453	0.348	*
					Other	0.000	-9.000	-9.000	
34	0-34	0.845	0.673	0.443	A	0.027	-0.291	-0.112	
					B	0.064	-0.488	-0.249	
					C	0.064	-0.648	-0.331	
					D	0.845	0.673	0.443	*
					Other	0.000	-9.000	-9.000	
35	0-35	0.827	0.482	0.326	A	0.036	-0.716	-0.305	
					B	0.827	0.482	0.326	*
					C	0.109	-0.402	-0.241	
					D	0.027	0.142	0.055	
					Other	0.000	-9.000	-9.000	

36	0-36	0.300	0.204	0.154	A	0.045	-0.357	-0.164
					B	0.027	-0.494	-0.190
					C	0.300	0.204	0.154 *
					D	0.627	-0.015	-0.012
					Other	0.000	-9.000	-9.000
37	0-37	0.100	0.259	0.151	A	0.664	-0.077	-0.060
					B	0.145	-0.070	-0.045
					C	0.091	-0.007	-0.004
					D	0.100	0.259	0.151 *
					Other	0.000	-9.000	-9.000
38	0-38	0.473	0.429	0.342	A	0.100	-0.527	-0.309
					B	0.473	0.429	0.342 *
					C	0.327	-0.071	-0.055
					D	0.100	-0.300	-0.175
					Other	0.000	-9.000	-9.000
39	0-39	0.064	-0.037	-0.019	A	0.664	0.117	0.090?
					B	0.227	-0.055	-0.039
		CHECK THE KEY			C	0.064	-0.037	-0.019 *
		C was specified, A works better			D	0.045	-0.224	-0.103
					Other	0.000	-9.000	-9.000

40	0-40	0.618	0.230	0.180	A	0.618	0.230	0.180	*
					B	0.136	-0.220	-0.140	
					C	0.073	-0.051	-0.027	
					D	0.173	-0.127	-0.086	
					Other	0.000	-9.000	-9.000	
41	0-41	0.209	0.433	0.306	A	0.291	-0.051	-0.039	
					B	0.355	-0.128	-0.099	
					C	0.145	-0.261	-0.169	
					D	0.209	0.433	0.306	*
					Other	0.000	-9.000	-9.000	
42	0-42	0.364	0.394	0.307	A	0.073	-0.616	-0.328	
					B	0.282	-0.023	-0.017	
					C	0.282	-0.163	-0.122	
					D	0.364	0.394	0.307	*
					Other	0.000	-9.000	-9.000	
43	0-43	0.436	0.344	0.273	A	0.364	-0.230	-0.179	
					B	0.436	0.344	0.273	*
					C	0.136	-0.278	-0.177	
					D	0.064	0.094	0.048	
					Other	0.000	-9.000	-9.000	

44	0-44	0.627	0.378	0.296	A	0.627	0.378	0.296	*
					B	0.173	-0.248	-0.168	
					C	0.136	-0.137	-0.087	
					D	0.064	-0.401	-0.205	
					Other	0.000	-9.000	-9.000	
45	0-45	0.691	0.615	0.469	A	0.064	-0.561	-0.287	
					B	0.145	-0.491	-0.318	
					C	0.691	0.615	0.469	*
					D	0.100	-0.197	-0.115	
					Other	0.000	-9.000	-9.000	

There were 110 examinees in the data file.

Scale Statistics

Scale: 0

N of Items	45
N of Examinees	110
Mean	20.364
Variance	25.086
Std. Dev.	5.009
Skew	0.313
Kurtosis	0.804
Minimum	6.000
Maximum	37.000
Median	20.000
Alpha	0.668
SEM	2.887
Mean P	0.453
Mean Item-Tot.	0.249
Mean Biserial	0.338

ANALYSIS OF VALIDITY

Using coefficient point biserial correlation:

r_p bis = $\frac{Xp - Xt}{St}\sqrt{p/q}$

item	n	true	false	p	q	p/q	$\sqrt{p/q}$	Xp	Xt	Xp-Xt	St	$r_p$bis	r table 5%	status
1	110	75	36	0,681818	0,318182	2,143	1,464	22,053	20,364	1,689	5,008	0,494	0,176	Valid
2	110	18	92	0,163636	0,836364	0,196	0,443	20,889	20,364	0,525	5,008	0,046	0,176	Invalid
3	110	54	56	0,490909	0,509090	0,964	0,982	21,648	20,364	1,284	5,008	0,252	0,176	Valid
4	110	50	60	0,454546	0,545454	0,833	0,913	22,4	20,364	2,036	5,008	0,371	0,176	Valid
5	110	5	105	0,045454	0,954545	0,048	0,219	23,6	20,364	3,236	5,008	0,142	0,176	Invalid
6	110	40	72	0,363636	0,636364	0,571	0,756	22,65	20,364	2.286	5,008	0,345	0,176	Valid
7	110	80	20	0,727273	0,272727	2,667	1,633	21,70	20,364	1,336	5,008	0,436	0,176	Valid
8	110	76	34	0,690909	0,309091	2,235	1,495	21,197	20,364	0,833	5,008	0,249	0,176	Valid
9	110	63	47	0,572727	0,427273	1,340	1,158	21,762	20,364	1,398	5,008	0,323	0,176	Valid
10	110	74	37	0,672727	0,327273	2,056	1,434	21.527	20,364	1,163	5,008	0,333	0,176	Valid
11	110	50	60	0,454545	0,545455	0,833	0,913	21,42	20,364	1,056	5,008	0,193	0,176	Valid
12	110	6	104	0,054545	0,945455	0,058	0,241	23,667	20,364	3,303	5,008	0,159	0,176	Invalid
13	110	13	97	0,118182	0,881818	0,134	0,366	22,846	20,364	2,482	5,008	0,181	0,176	Valid
14	110	63	46	0,572727	0,427273	1,340	1,158	21,143	20,364	0,779	5,008	0,180	0,176	Valid
15	110	21	89	0,190909	0,809091	0,236	0,486	22,238	20,364	1,874	5,008	0,182	0,176	Valid
16	110	27	83	0,245455	0,754545	0,325	0,570	23,518	20,364	3,154	5,008	0,359	0,176	Valid
17	110	78	32	0,709091	0,290909	2,438	1,561	21,359	20,364	0,995	5,008	0,310	0,176	Valid
18	110	84	26	0,763636	0,236364	3,230	1,797	21,583	20,364	1,219	5,008	0,437	0,176	Valid
19	110	42	68	0,381818	0,618182	0,618	0,786	22,786	20,364	2,422	5,008	0,380	0,176	Valid
20	110	37	73	0,336364	0,663636	0,507	0,712	22,892	20,364	2,528	5,008	0,359	0,176	Valid
21	110	39	71	0,354545	0,645455	0,549	0,741	19,692	20,364	-0,672	5,008	-0,099	0,176	Invalid
22	110	34	76	0,309091	0,690909	0,447	0,669	20,941	20,364	0,577	5,008	0,077	0,176	Invalid
23	110	40	69	0,363636	0,636364	0,571	0,756	21,35	20,364	0,986	5,008	0,149	0,176	Invalid
24	110	73	36	0,663636	0,336364	1,973	1,405	21,219	20,364	0,855	5,008	0,239	0,176	Valid
25	110	86	24	0,781818	0,218182	3,583	1,893	19,686	20,364	-0,678	5,008	-0,256	0,176	Invalid
26	110	73	36	0,663636	0,336364	1,973	1,405	21,041	20,364	0,677	5,008	0,189	0,176	Valid
27	110	31	79	0,281818	0,718182	0,392	0,626	23,032	20,364	2,668	5,008	0,334	0,176	Valid
28	110	22	88	0,2	0,8	0,25	0,5	23,045	20,364	2,681	5,008	0,268	0,176	Valid

29	110	54	56	0,490909	0,509091	0,964	0,982	22,111	20,364	1,747	5,008	0,343	0,176	Valid
30	110	86	22	0,781818	0,218182	3,583	1,893	21,012	20,364	0,648	5,008	0,245	0,176	Valid
31	110	46	74	0,418182	0,581818	0,719	0,848	21,283	20,364	0,919	5,008	0,156	0,176	Invalid
32	110	15	95	0,136364	0,863636	0,158	0,397	23	20,364	2,636	5,008	0,209	0,176	Valid
33	110	74	36	0,672727	0,327273	2,056	1,434	21,581	20,364	1,217	5,008	0,348	0,176	Valid
34	110	93	17	0,845455	0,154545	5,471	2,339	21,312	20,364	0,948	5,008	0,443	0,176	Valid
35	110	91	19	0,827273	0,172727	4,789	2,188	21,109	20,364	0,745	5,008	0,325	0,176	Valid
36	110	32	77	0,290909	0,709091	0,410	0,640	21,625	20,364	1,261	5,008	0,161	0,176	Invalid
37	110	12	98	0,109091	0,890909	0,122	0,349	22,333	20,364	1,969	5,008	0,137	0,176	Invalid
38	110	52	58	0,472727	0,527273	0,897	0,947	22,173	20,364	1,809	5,008	0,342	0,176	Valid
39	110	7	103	0,063636	0,936364	0,068	0,260	20	20,364	-0,364	5,008	-0,019	0,176	Invalid
40	110	68	42	0,618182	0,381818	1,619	1,272	21,073	20,364	0,709	5,008	0,180	0,176	Valid
41	110	23	87	0,209091	0,790909	0,264	0,514	23,348	20,364	2,984	5,008	0,306	0,176	Valid
42	110	40	70	0,363636	0,636364	0,571	0,756	22,4	20,364	2,036	5,008	0,307	0,176	Valid
43	110	48	62	0,436364	0,563636	0,774	0,879	21,917	20,364	1,553	5,008	0,273	0,176	Valid
44	110	69	41	0,627273	0,372727	1,683	1,297	21,507	20,364	1,143	5,008	0,296	0,176	Valid
45	110	76	34	0,690909	0,309091	2,235	1,495	21,934	20,364	1,57	5,008	0,469	0,176	Valid

ANALYSIS OF VALIDITY

Using coefficient point biserial correlation:

r_p bis = $\frac{Xp - Xt}{St}\sqrt{p/q}$

item	n	true	false	p	q	p.q	∑p.q	Xp	Xt	Xp-Xt	St	$r_p$bis	r table 5%	status
1	110	75	36	0,681818	0,318182	0,216942	8,658	22,053	20,364	1,689	5,008			
2	110	18	92	0,163636	0,836364	0,136859	8,658	20,889	20,364	0,525	5,008			
3	110	54	56	0,490909	0,509090	0,249917	8,658	21,648	20,364	1,284	5,008			
4	110	50	60	0,454546	0,545454	0,247934	8,658	22,4	20,364	2,036	5,008			
5	110	5	105	0,045454	0,954545	0,043388	8,658	23,6	20,364	3,236	5,008			
6	110	40	72	0,363636	0,636364	0,231404	8,658	22,65	20,364	2.286	5,008			
7	110	80	20	0,727273	0,272727	0,198347	8,658	21,70	20,364	1,336	5,008			
8	110	76	34	0,690909	0,309091	0,213554	8,658	21,197	20,364	0,833	5,008			
9	110	63	47	0,572727	0,427273	0,244710	8,658	21,762	20,364	1,398	5,008			
10	110	74	37	0,672727	0,327273	0,220165	8,658	21.527	20,364	1,163	5,008			
11	110	50	60	0,454545	0,545455	0,247934	8,658	21,42	20,364	1,056	5,008			
12	110	6	104	0,054545	0,945455	0,051569	8,658	23,667	20,364	3,303	5,008			
13	110	13	97	0,118182	0,881818	0,104215	8,658	22,846	20,364	2,482	5,008			
14	110	63	46	0,572727	0,427273	0,244710	8,658	21,143	20,364	0,779	5,008			
15	110	21	89	0,190909	0,809091	0,154463	8,658	22,238	20,364	1,874	5,008			
16	110	27	83	0,245455	0,754545	0,185207	8,658	23,518	20,364	3,154	5,008			
17	110	78	32	0,709091	0,290909	0,206281	8,658	21,359	20,364	0,995	5,008			
18	110	84	26	0,763636	0,236364	0,180496	8,658	21,583	20,364	1,219	5,008			
19	110	42	68	0,381818	0,618182	0,236033	8,658	22,786	20,364	2,422	5,008			
20	110	37	73	0,336364	0,663636	0,223223	8,658	22,892	20,364	2,528	5,008			
21	110	39	71	0,354545	0,645455	0,228843	8,658	19,692	20,364	-0,672	5,008			
22	110	34	76	0,309091	0,690909	0,213554	8,658	20,941	20,364	0,577	5,008			
23	110	40	69	0,363636	0,636367	0,231406	8,658	21,35	20,364	0,986	5,008			
24	110	73	36	0,663636	0,336364	0,223223	8,658	21,219	20,364	0,855	5,008			
25	110	86	24	0,781818	0,218182	0,170579	8,658	19,686	20,364	-0,678	5,008			

26	110	73	36	0,663636	0,336364	0,223223	8,658	21,041	20,364	0,677	5,008			
27	110	31	79	0,281818	0,718182	0,202397	8,658	23,032	20,364	2,668	5,008			
28	110	22	88	0,2	0,8	0,16	8,658	23,045	20,364	2,681	5,008			
29	110	54	56	0,490909	0,509091	0,249917	8,658	22,111	20,364	1,747	5,008			
30	110	86	22	0,781818	0,218182	0,170579	8,658	21,012	20,364	0,648	5,008			
31	110	46	74	0,418182	0,581818	0,243305	8,658	21,283	20,364	0,919	5,008			
32	110	15	95	0,136364	0,863636	0,117769	8,658	23	20,364	2,636	5,008			
33	110	74	36	0,672727	0,327273	0,233802	8,658	21,581	20,364	1,217	5,008			
34	110	93	17	0,845455	0,154545	0,130661	8,658	21,312	20,364	0,948	5,008			
35	110	91	19	0,827273	0,172727	0,142892	8,658	21,109	20,364	0,745	5,008			
36	110	32	77	0,290909	0,709091	0,206281	8,658	21,625	20,364	1,261	5,008			
37	110	12	98	0,109091	0,890909	0,097190	8,658	22,333	20,364	1,969	5,008			
38	110	52	58	0,472727	0,527273	0,249256	8,658	22,173	20,364	1,809	5,008			
39	110	7	103	0,063636	0,936364	0,059586	8,658	20	20,364	-0,364	5,008			
40	110	68	42	0,618182	0,381818	0,236033	8,658	21,073	20,364	0,709	5,008			
41	110	23	87	0,209091	0,790909	0,165372	8,658	23,348	20,364	2,984	5,008			
42	110	40	70	0,363636	0,636364	0,231405	8,658	22,4	20,364	2,036	5,008			
43	110	48	62	0,436364	0,563636	0,245950	8,658	21,917	20,364	1,553	5,008			
44	110	69	41	0,627273	0,372727	0,233802	8,658	21,507	20,364	1,143	5,008			
45	110	76	34	0,690909	0,309091	0,213554	8,658	21,934	20,364	1,57	5,008			

APPENDIX D

STUDENTS' ANSWER DISTRIBUTION

045 o N 03

BCCADBCDDAADCCBCADDCCBBACADDBCAADDBCDBCADDBAC

444444444444444444444444444444444444444444444

YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

001BBCACACBAAABBDCCADDCADCACBBCBCCDDDBCACBADDBAC

002BBAACBCDDADBCBCBADBCBBDCCBADABBCDDBDACBABCDAC

003BAAACDCDCBABBCAAADAADDBACACABDBCBCBDAACDBCBAC

004BBDACCCBCDABBDCAADCDABDACACCACBCDDBCAAAABCBAC

005BBCDCCCDDACABDDBAABDBDBCCACDBCADBDCABBDCACBDB

006BBADCACDDADBBDCAADACCDCACABDADBBDDBCACBABDAAC

007CABDACADBCADBDBACDDACBCABDBCADCBDACADBCABDADB

008CCDDCCADADABAACDADCDCDCACACCCDDCACBDACADBCCAC

009DADBCBCBCADABDBBBDDACAAACAACBCACCBCDDDBABDAAB

010BCBCBCBDBABBABBBCDACBCBACBCCBCABCBCDCBABBCBCB

011BABACCCDBAABBBBABDBABBAACBCCBCCBADBDABAABBACA

012BBDACBDDAADBBACCADBBDBCBCADABCCCDDBBAAAABDAAC

013DACACBCBBADBBBCBADCCDDBCCABDACCBDACDACADBBCAA

014ABBCCDBBDBDBBAAACCBDACDBDADABCDDBDBCCBBBCABDB

015BCDACBBDDAABBDDCADDCDDDACBCDBCBADDBCABBACDCAC

016CCDBCCCDBADBBDDAADBCBDBACBCBCCBBCDBACBACCCAAC

017CAACCCDDDCBBCBABADCABBBACAACBCACCCADDDAABBAAB

018BAABCCCBCAABBCCABBBDCBCCDDCCACACDDBCACAAACAAC

019ADCACBBDBCDACCAABADBCBACBDCACCABCBBDAAABCBBBD

020BBDCCACBDBABBCBACDCACBCACACCBCCCADCDADAAACABD

021BBCACBCDDDDBBCAADBAABDBACAABBCCDDDBAABAABCBBC

022BBCADACDDDABBCABADDADDCAAACCBCBDDDCCABAAADCBC

023BBBDCBCDDAABBCDAADDCCBCACADBADCDDDBDACADBCBAC

024BCCDCCCDDABCBCAABDBADDCACCCCACABDDBDACABABAAC

025BACDCCCCCABBBCAABBAABBCACBDCBCBADDBDABADACBAC

026CBCACCDBDADBCADAADBABCBACAACCCBCDBCCDCAADCACB

027BCCACCCDDADBBCBABDBDCDDACCCCACCCDDBDACAACDAAC

028CBDDCCDCACBABAAAABCACADACBCADACCACABAABACACAA

029CBAACBDDDADBBACAAACBBBCACAACCCCCBDACCCBCBCAAD

030BACCCDCDAADBBDCABDDCCBDACADAACBCADBDDBADDDBAC

031DAACCCCDCBAABCCAAABDCABCCAACACABADBCBDBDBDAAC

032BCCDCCCDAAABBBAAADDCABCACACDBCCDDDBDBBAABCACC

033DCAACACBDAABBCAAADCDBDBCCBBCBDCCDDBCBAADACBCB

034DCAACACBBCBABBBAACDDDACACDCCDBCCBDBDBACBACBCB

035ABCDCCCBDADABBBAADBABCBACACCACACDDBDADADCDCAC

036DADDCCDDBBABBCAADDBABCBCCABCBCACCCCAADAAAABAC

037BBCCCBCCDAABCCCAADADCDBCCDAADCACDADCACCBCBDAC

038DACDCCCDDADABCCCAABDDDDACAACCCBCDDBDACAACACBC

039BBDDCCCDAADBBADDADDCDCBACADDACBADDBDACAADCBAC

040DAAACCCBBADABCBCBDCACCAACACDCCBCBDBDABBAABBAB

041CAAACCADBAABBACBAADCCBCCBACABCDCABBDABADABCBD

042BBDDCBCDDAABCCBCADDCCBCABBDABCACBDBDDBBADDBAC

043CABCCCCDAAABBCCAADDBCCDCCACDCAACDDBDBCAADDDAA

044ABDDCBDDCADBBCCDADAACCDACBDBBCBCBCBCACAACCACC

045BBBDCBCDDADBCABCADDCDBCABADBDCBCDDBCDBADADAAC

046BBCADCCDDDABBCBCADDBDBBABADDBCBADDBCBCBABDBCD

047BBCDCACDDDABBCABADCABDCBDBCBBCADDCBDACDBCDACC

048BBCADBCDDADABABBADDDCDAACACABCACDDBDABDBABAAC

049BBCACCCDDAABBBBDADDBDCDCCACBACCCDDBDACABACAAC

050BBDDCCCDDADCBCDAACDCCCBACBAAACCAADCDBCBDABBAC

051BCCACCDDDAAABABAADCCBDBCCAABBCABDDBDABAAABCAC

052BBCACBCBAAABBABADDDCBCBABACCBCACDDBDADAADABAC

053BBCACCCDDBDABCAAADCCCCBBCACBADDABDBDACAAABBAC

054BACACCCDDBDDBCAAADCBCCBBCACBADDABDBDACAABDBBB

055BACDCCCDDBDDBCAAADCBCCBBCACBADDABDBDACAABDBBB

056BACACBCDCDABBCACBBADACDACAACADCABDBDABAABDAAC

057BBBACBCDDAADCCBDADDCDDBABADDBCAADDBCCBCADDBAC

058BCCACBCDDADBCCCCADDCDBDABADDBCDBDDBDBBACADBAC

059BACDCACDCADBBADDADCAABBACADACCACDDBDABADADBAB

060CCCBDCBAACDBCADDDABACBDCCABDCAACDDBCBADBBCADB

061CBCACACCBADBBBABACDDCDDCCAAABCAADDBCCBAADBBAC

062BBCDDBCADBABBCCCADCCCDBCCACABCBBADBDABACDBBAC

063BBCDCBCDDADABCCCADBADCBAAACCBCCCADBDABABDBBAC

064BBCDCCDDCADBBAACADDDCDAACABCBCCBDDBCABADADAAC

065BBDACBCBBADBBDCAADCDBDBACACBCCABDDBCDBAABDBAC

066BCBACCDBBAABBAABADDCDDCACAADACDDDDBDABBCAABAC

067BBCACBCDDAABBBCBADDADDBACBDCBCBCDDBCACAADBBAC

068CADDCCDBDDABBADAACADCDDACBBACCCCBDBDBBAAAAABC

069BBCDBBBDCBDBBCCCADCBADCBAACDCDBCADBDBDAABBBAD

070BCCACBCDDADBBCCAADCCBCBACABABCADDDBDABAAACDCC

071BBAACBCDBAAABACCADBAADCACCBCDCABDDBDAAAACCAAC

072BBDBCBDBCAABBDBBCADDCAABCBDBACBCDBBCDABCBDBDB

073BBBACBCDDAAABCDCADDBBACBCADABCCABDBCDBABADBAC

074BBCACCCDDADBCADDDDAABDCACBDABCCCDDBDACAADDBAC

075CBBBCACDDADBBBDDDDCBDCCCCBCDBCCBDDBCABAADCAAC

076BDCBCCCDCADABDCABBBCDAACCADCACBADDBCBBADDDADB

077CBDDCBCDDAABBCDCAABDBCCACABAACBBCDBDABADDBBAC

078BBCACBCDDAADDCBCADBCCDBAAACDBCACDDBCDCAABCBAC

079CADACACDDBDBBCDAADDCBDCCCADDBCCCDDBDCBAABBBAC

080ABCBCBADBCADBBDACADBBCBACADCACABDDBDACBADCABD

081BBCACCCDDAABBCBDADACBDCAAACBACABDDBDCBBABDCAC

082DACACACDDCDBBCCAADADDBDCCADBBDACADBCCBAADBACC

083CCCDCCCBDAACBCCAADBDCCACCACDBCBDDDBCACAAABAAC

084BBCDCBDDBADBBCCDADADCBCACADBBDACDDBDABAADDACD

085BBCDCCDCDDABBCCABDADDBCACADBBDACDDBDABAADDACD

086BBCDCCDBDABABCBBBDDADDDABABABCACDDBDABAABCDBC

087BBBDCCDCBABABCDABDBCDCDACCCBCDBDDDBDACBAAACBA

088BBDCCCCDDABABCDABDBCBACACBCBACBADDBCCDBDADAAC

089BBDACCCDDADBBCCCADBCBCDACACBBCACDDBDABAADBAAC

090BBDDCCCDDDDBBCDCADBDCDBCCBDAADBCDDBDACCCDBAAC

091DAABCBCDACDABCDABABCDCBACADDACACDDCDACCDBBAAC

092DBCDCCCDDAABBCDDCAADCDBBCACAACBCDDBBAADAABBCA

093BBCDCBCDDADBBCACADBBDCDACBDBBDACDDBDABAABDDBC

094BABBCADDCBDBBCAAAACBACDCCAABBCABDDACADAABBCBC

095BBCDCCCDBADABCACBDDDBDDAAABCACABDDBDABAABDDBC

096BBCACDCBDABBBCACADDCDBDACADCACBDDDBDBDAABBCBC

097BCCDCCDACBDBBCACADDCBACDCACACCACDDBDBCABADBCA

098BBBBCCCDDCABBDABADACBBDACDCAADACDBCDACBABDAAC

099BCCACBCDABDBBCADADDCBDBABADDBCACDDBCBBBADCBAC

100BCCDCCCDCAABBCDBADACBDCACCCDBCACDDBDABAAADBDC

101BADACCCDDBABBCDBADDDCBBCCACAACACCDDDABABABBAC

102BBDDCBCDDCDABCDBADDCADCACBDABCACDDBDABABABBAC

103BBCABCCCBAABCCDCADDDDBAAABDABCACDDBDABAACCABC

104CBDDCBCBDAABBCCAAACDBBBAAABDBCBCDDBDACAABDBAC

105DBBDCBDBDAABBCDBAACABDBACADABCBAADBDABBABCCAC

106BBDACBCBDAAABDDABDAABBBACBCBACBDDDBCACAACDBAC

107BBDACBADDAAABCACADDACBAABACAAAACDDBDABADBDBAC

108DABACCCCDABBBCDADDDDCBBDCBCABCABDDBDABBACBCAD

109DACACCCDDAABCCDCADDDCBACABCCBCBCDDBCACBABCABD

110BBDACCCDDAABBCDBADDACBBACACACCACCDDDABBACBABC

APPENDIX E

SUMMARY OF THE RESULTS OF THE ANALYSIS OF EACH TEST ITEM OF THE FIRST TERM TEST

No	Validity index	Difficulty index	Discrimination index	Effective distractor	Interpret				Explanation
					Validity	Difficulty index	Discriminati on index	Distractor	
1	0,494	0,673	0,639	A,B,C,D	Valid	Medium	Very good	Excellent	Good
2	0,046	0,164	0,070	A,B,C	Invalid	Difficult	Poor	Good enough	Poor
3	0,252	0,491	0,316	A,B,C,D	Valid	Medium	Good	Excellent	Good
4	0,371	0,455	0,466	A,B,C,D	Valid	Medium	Very good	Excellent	Good
5	0,142	0,045	0,307	B,C,D	Invalid	Difficult	Good	Good enough	Poor
6	0,345	0,355	0,462	A,B,C,D	Valid	Medium	Very good	Excellent	Good
7	0,436	0,727	0,584	A,B,C,D	Valid	Easy	Very good	Excellent	Good enough
8	0,249	0,691	0,326	A,B,C,D	Valid	Medium	Good	Excellent	Good enough
9	0,323	0,573	0,408	A,B,C,D	Valid	Medium	Very good	Excellent	Good
10	0,333	0,673	0,433	A,B,C,D	Valid	Medium	Very good	Excellent	Good
11	0,193	0,455	0,242	A,B,D	Valid	Medium	Marginal	Good enough	Poor
12	0,159	0,055	0,325	A,B,C,D	Invalid	Difficult	Good	Excellent	Poor
13	0,181	0,118	0,296	B,C	Valid	Medium	Marginal	Not good	Poor
14	0,180	0,582	0,273	A,B,C,D	Valid	Medium	Marginal	Excellent	Good enough
15	0,182	0,191	0,263	A,B,C,D	Valid	Difficult	Marginal	Excellent	Poor
16	0,359	0,245	0,491	A,B,C,D	Valid	Difficult	Very good	Excellent	Good enough
17	0,310	0,709	0,411	A,B,C,D	Valid	Easy	Very good	Excellent	Good enough
18	0,437	0,764	0,603	A,B,C,D	Valid	Easy	Very good	Excellent	Good enough
19	0,380	0,382	0,484	A,B,C,D	Valid	Medium	Very good	Excellent	Good
20	0,359	0,336	0,465	A,B,C,D	Valid	Medium	Very good	Excellent	Good
21	-0,099	0,355	-0,128	A,B,C,D	Invalid	Medium	Poor	Excellent	Poor
22	0,077	0,309	0,101	A,B,C,D	Invalid	Medium	Poor	Excellent	Poor

23	0,149	0,364	0,191	A,B,C,D	Invalid	Medium	Poor	Excellent	Poor
24	0,239	0,664	0,311	A,B,C	Valid	Medium	Good	Good enough	Poor
25	-0,256	0,782	-0,359	A,B,C,D	Invalid	Easy	Poor	Excellent	Poor
26	0,189	0,664	0,246	A,B,C,D	Valid	Medium	Marginal	Excellent	Good enough
27	0,334	0,282	0,445	A,B,C,D	Valid	Difficult	Very good	Excellent	Good enough
28	0,268	0,209	0,370	A,B,C,D	Valid	Difficult	Good	Excellent	Poor
29	0,343	0,491	0,429	A,B,C,D	Valid	Medium	Very good	Excellent	Good
30	0,245	0,782	0,343	A,C,D	Valid	Easy	Good	Good enough	Poor
31	0,156	0,418	0,196	A,B,C,D	Invalid	Medium	Poor	Excellent	Poor
32	0,209	0,136	0,328	A,B,C,D	Valid	Difficult	Good	Excellent	Poor
33	0,348	0,673	0,453	A,B,C,D	Valid	Medium	Very good	Excellent	Good
34	0,443	0,845	0,673	A,B,C,D	Valid	Easy	Very good	Excellent	Good enough
35	0,325	0,827	0,482	A,B,C,D	Valid	Easy	Very good	Excellent	Good enough
36	0,161	0,300	0,204	A,B,C,D	Invalid	Difficult	Marginal	Excellent	Poor
37	0,137	0,100	0,259	A,B,C,D	Invalid	Difficult	Marginal	Excellent	Poor
38	0,342	0,473	0,429	A,B,C,D	Valid	Medium	Very good	Excellent	Good
39	-0,019	0,064	-0,037	A,B,C,D	Invalid	Difficult	Poor	Excellent	Poor
40	0,180	0,618	0,230	A,B,C,D	Valid	Easy	Marginal	Excellent	Poor
41	0,306	0,209	0,433	A,B,C,D	Valid	Difficult	Very good	Excellent	Good enough
42	0,307	0,364	0,394	A,B,C,D	Valid	Medium	Good	Excellent	Good enough
43	0,273	0,436	0,344	A,B,C,D	Valid	Medium	Good	Excellent	Good enough
44	0,296	0,627	0,378	A,B,C,D	Valid	Medium	Good	Excellent	Good enough
45	0,469	0,691	0,615	A,B,C,D	Valid	Medium	Very good	Excellent	Good